International Journal of American Linguistics

VOLUME XXI April, 1955 NUMBER 2

LINGUISTIC TIME DEPTH RESULTS SO FAR AND THEIR MEANING

A. L. KROEBER

UNIVERSITY OF CALIFORNIA, BERKELEY

0. Glottochronology and other linguistic estimates of time
1. Western and Eastern Eskimo; Aleut and proto-Eskimo
2. Tlingit and 'Hupa'; the composite nature of the 'Hupa' data
3. Salish: Lower Fraser and Lillooet; Pentlatch and Kalispel . . . Uto-Aztecan: Papago and Paiute; Nahua and Papago; Nahua and Paiute . . . Wakashan: Nootka and Kwakiutl . . . Mayan: Yucatec and Huastec
4. Hokan: Chimariko and Yana; YANA and Yuman and Chontal and Jicaque and Comecrudo and Washo; YUMA and Chontal and Jicaque and Comecrudo and Washo; CHONTAL and Jicaque and Comecrudo and Washo; JICAQUE and Comecrudo and Washo; Comecrudo (Coahuiltecan) and Washo; Atakapa and Chitimasha
5. Penutian: YOKUTS and Chinook and Tsimshian; Chinook and Tsimshian . . . compare in general: Uto-Aztecan; Pacific Hokan; other Hokan
6. TOTONAC and Zoque and Huave and Yucatec; YUCATEC and Zoque and Huave; Zoque and Huave
7. Mosan: proto-Salishan and Quileute; Kwakiutl and Quileute
8. Internal time distances in Eskimo, Athabascan, Uto-Aztecan, Pacific Hokan, Macro-Penutian, Huave-Mayan compared with those in Salish: South Georgian; South Georgian and Puget; South Georgian and North Georgian; SOUTH GEORGIAN and Twana and Bella Coola and Olympic and Tillamook; South Georgian and Interior Salish
9. Alternative interpretations of the Salish anomaly: the frontiers of the Salish linguistic area
10. Continuity in geographic distribution of Salish contrasted with discontinuities in Pacific Hokan: (1) Chimariko, Karok, Shasta, Achomawi, Yana; (2) Pomo; (3) Esselen, Salinan, Chumash; (4) Yuman block; (5) Seri; (6) Tlapanec; (7) Chontal; (8) Jicaque; (9) Subtiaba
11. Remaining problems in the perspective of relatable sets of phenomena

0. In the American Anthropologist for 1953, Quimby, Collins, Haury, Ekholm, and Eggan have commented on a brief paper in the same issue by Swadesh.[1] They consider the fit between Swadesh's glottochronological results there summarized, and archaeological and ethnological age estimates arrived at in other ways.

I shall here compare the same glottochronological results for their fit and consistency with other LINGUISTIC facts and estimates.

It is of course true that the glottochronological method assumes for the time being that all or most languages, whatever their kind or type, change at a pretty uniform, steady rate. This assumption has not been proved; but it is only its application to as many instances as possible that will prove, disprove, or modify it. For this purpose the crucial cases will be families in which there occurred ancient, usually now extinct languages preserved in written form, such as in the Semitic, Hamitic, Indo-European families. However, enough is known with

[1] Time Depths of American Linguistic Groupings, AA 56.361-364 (1954).

reasonable accuracy of many American Indian languages to make it interesting to match the results of ordinary 'inspectional' comparison, and geographical positions, against the Swadesh quantitative-sample results. Off and on I have been concerned for fifty years with the inspectional comparison for historical purposes and mapping of many of these languages.

A practical approach seems to be to proceed from the more similar languages of recent divergence (high percentages of cognates and relatively few centuries since the split) to the more differentiated languages that have been separate for many centuries. 'C' stands for centuries of divergence, '%' for percentage of cognates retained.

1. Western and Eastern Eskimo (Yukon-Greenland, 10 C, 66 %) show the most recent separation of the languages reported on by Swadesh. This is as it should be, for the close relationship of Alaskan Eskimo to Greenlandic has been evident ever since southwestern Eskimo vocabularies were available.

Swadesh's asterisked figure for Eskimoan-Aleut 'Eskaleut' stock—30 C, 36 %—is non-comparable with intra-Eskimo Yukon-Greenland 10 C, because it matches reconstructed proto-Eskimo against contemporary Aleut.

2. The next lowest of Swadesh's figures is 20 C, 44 %, for 'Tlingit-Hupa etc.' This would mean, on its face, that Tlingit is as similar to Hupa taken as a representative of Athabascan as Comox is to near-by Pentlatch (20 C, 45 %) within the North Georgian branch of Coast Salish. This is obviously hardly so, else Tlingit would have been classed in Athabascan (or 'Dene') as soon as that stock was set up, instead of their kinship awaiting Sapir's much later discovery. It is evident from Swadesh's list as a whole that a relationship only 20 or even 30 centuries old is soundly discoverable by inspection, provided vocabularies are fair phonetically, are more than mere snatches, are reasonably well chosen for spread, and that the comparer has no marked bias for the fantastic.

The trouble in this instance is evidently in the 'Hupa etc.,' namely the fact that the data go back to the old Goddard and Boas controversy with Sapir. Goddard,[2] in trying to show that Tlingit could not be proved to be Athabascan, assembled words from half a dozen different Athabascan languages showing similarity to Tlingit, with the idea that even this compound mustering was insufficient, and would backfire to establish his case for relationship being not provable. In this list were 82 forms whose meanings are in Swadesh's basic vocabulary of 200; and of these 82, Swadesh finds 36 to be cognate, or 44 %.[3] However, this composite basis of comparison is different from Swadesh's usual language-to-language comparison: it is language-to-several-languages. The date in centuries therefore means something else, and should not be used.

Between what Sapir, Li, Hoijer and others have recorded, it ought not be difficult to muster adequate basic vocabularies for a dozen Athabascan languages with satisfactory geographic spread. After that, comparison with Tlingit would carry more meaning, because it would show whether Tlingit was specially connected with one Athabascan subdivision, about equally with two or more, or more remotely but substantially evenly with all of them, which would make it coordinate within 'Dene' with proto-Athabascan.

In spite of its wide geographic spread it is obvious that Athabascan has well-marked

[2] IJAL 1.266–279 (1920).

[3] As a matter of fact, in his original article, Diffusional Cumulation and Archaic Residue, in SJA 7.14 (1951), Swadesh says that Tlingit and Athabascan "separated more than 2000 and less than 5000 years ago." It is evident that overcondensation has led to inaccuracy of statement in A.A. 1954; '20–50 C' is undoubtedly more in accord with all his other results than the abbreviated 20 C.

consistency; and yet I should expect at least 20 centuries of separation, quite probably up to fully 30, for some of the more divergent of its branches, rather than for Tlingit from Athabascan.

3. Next come splits around 3,000 years old. Here we have Lower Fraser of Coast Salish and Lillooet of Interior Salish (30 C, 28%)—cited no doubt because their territories are adjacent, since other Coast-Interior Salish languages, such as Pentlatch and Kalispel, are separated by as much as 50 C, 10%. The problem raised by this difference is gone into below.

In this same range Swadesh also puts three Uto-Aztecan languages: Papago-Paiute (presumably Southern Paiute: 27 C, 32%); Nahua-Papago (29 C, 29%); Nahua-Paiute (31 C, 27%). While the reconstruction of original Uto-Aztecan was not begun until 1913 by Sapir, the name was coined and the kinship asserted by Brinton in 1891.[4] In this I concurred, and gave further evidence, in Shoshonean Dialects in 1907.[5] In fact Buschmann as early as 1859 in his Spuren der Aztekischen Sprache[6] had pointed out Aztec and Sonoran resemblances to various Shoshonean languages, though he seems to have construed the similarities as due to contact influences and not to common descent: his difficult German is none too clear, but he certainly did NOT set up a GENETIC family.[7] Powell's linguistic classification for North America, though contemporaneous with Brinton's, also did not accept a common stock, but kept Shoshonean separate from Piman (Sonoran) and from Aztec; though Powell was aware that Gatschet and Brinton did unite all three. Powell was evidently influenced by a bias in favor of particularity for the Ute-Southern Paiute vocabularies he had personally collected in his geological travels.

In short, we have in Uto-Aztecan a group of languages whose genetic relationship has been generally recognized (except for some formal doubts) on inspectional comparison for over sixty years, was confirmed by Sapir on a reconstructive basis in 1913–15,[8] (and extended by Whorf and Trager into Azteco-Tanoan[9]); and we now have three geographically well-spaced members of Uto-Aztecan indicated as diverging from one another around 3000 years ago.

This figure may prove generally significant for languages classed as related within the same Powellian 'stock' or 'family'—that is, on systematic comparison, by linguistically interested non-linguists, of a moderately sized, reasonably chosen vocabulary recorded with some modest experience in hearing and writing foreign idioms.

That it is to somewhere in this vicinity of 3000 years of divergence that commonsense nonreconstructive comparison can penetrate is shown by Swadesh's Nootka-Kwakiutl (Powell's Wakashan, 29 C, 30%); and Yucatec-Huastec (Mayan, 32 C, 26%).

It ought now be not at all difficult to rough out the history of internal relations in both Mayan and Uto-Aztecan. The Stoll vocabularies from Guatemala are presumably not quite long enough, but they provide an excellent start, and with the materials of Andrade, McQuown, and others, lists of close to 200 vocables should be rather readily providable. In 1939 I attempted a Mayan classification[10] which would go beyond Stoll's[11] and be free of some patent unsoundness of Gates's.[12] This

[4] The American Race, 1891.

[5] UC-PAAE 4.65–165, 1907.

[6] Abh. Kön. Akad. Wiss. Berlin 1854, 2nd supplem. vol., 1859.

[7] See a passage quoted from him in my Shoshonean Dialects, p. 156.

[8] S. Paiute and Nahuatl: A Study in Uto-Aztecan. I, JSAP 10.379–425 (1913); II, AA 17.98–120, 306–328 (1915).

[9] AA 39.609–624 (1937).

[10] Cult. and Nat. Areas, UC-PAAE, 38.114 (1939).

[11] Zur Ethnographie der Republik Guatemala, 1884.

[12] Appendix XII of S. G. Morley, Inscriptions at Copan, Carn. Inst. Publ. 219, 1920.

provisional classification seems to have been overlooked both by linguists and by Mayan specialists, through being buried in a book on North American culture, but could now be rather readily tested and corrected by glottochronology.

Uto-Aztecan grand classification is certainly determinable from data already recorded, especially from the United States. I shall be glad to contribute Luiseño as representative of the South Californian branch; Voegelin, no doubt Tübatulabal; Shimkin or Casagrande are available for Shoshoni-Comanche (the latter's collecting specialized on semantics!); if no living worker has Northern Paiute data, Marsden's manuscript materials[13] might yield a reasonably adequate list; for Hopi, Kennard undoubtedly has a vocabulary if Whorf did not pause to collect a basic one; and the notebooks of Mason and the late Jean Johnson almost certainly can provide Yaqui plus another Taracahitian language; only Huichol and Cora have been visited by ethnologists rather than linguists in recent decades. On top of this all, Trager's materials will give a parallax on the remoter forking between Uto-Aztecan and Tanoan—corresponding say to that between Indo-European and Hittitte.

I might add that instrumental verb prefixes are conspicuous in all three branches of Basin-Plateau Shoshonean (Northern and Southern Paiute, Shoshone), but seem to be totally lacking from every other Uto-Aztecan language whose structure is known.[14] If these prefixes are confirmed as limited to Basin-Plateau Shoshonean, they must represent either an autonomous structural growth in proto-Plateau Shoshonean or perhaps more likely a development due to stimulus from outside, such as from Siouan. With the approximate age of Plateau-Shoshonean determined by glottochronology, we might then be in position to reconstruct an inference on what division of Siouan (or wider stock) had been the influencer; and, with enough internal evidence on that stock, the approximate region in which the influencing occurred might be indicated.

4. Let us now climb another rung down the ladder of descent—from around 30 centuries ago to the period of 35 to 39. Here there are some intra-Salish relations—which I would like to reserve for discussion below—and some within Hokan, one of the superstocks first recognized by Dixon and Kroeber in 1913–19,[15] though later enlarged by Sapir and others into Hokan-Siouan, etc.

For Hokan, Swadesh cites seven coefficients for seven languages in his 1954 article. Five of these coefficients fall between 35 and 39 C, one is 44, and one 55. We cannot however infer that Hokan separations mostly are less than 4000 years old, because Swadesh's citations are selective. Seven languages of course would yield a total of 21 coefficients. It proves that six of his seven coefficients are taken from a 1953 article by Greenberg and Swadesh on Jicaque,[16] where the full 15 cognate percentages of six Hokan languages are given. To these Swadesh has now added for Chimariko only its Yana relation. All 16 of these comparative figures are obviously required for full understanding of the situation so far as it has developed. I therefore present them retabulated (Table 1) (with the 1953 percentages converted into centuries as in the 1954 article), and in an order arranged to bring out certain new features.

Instead of five of seven time depths under

[13] Of which an excerpt appears in UC-PAAE, 20.175–191 (1923).

[14] With the possible exception of Papago, where Mason, The Papago Language, Univ. Penn. Mus. Monograph, 1950, p. 17, points out 3 elements in fossilized compounds: ki- *with teeth*, ma- *hand*, mu- *nose*.

[15] New Ling. Fam. in Calif., AA 15.647–655 (1913); Ling. Fam. of Calif., UC-PAAE 16.47–118 (1919).

[16] J. Greenberg and M. Swadesh, Jicaque as a Hokan Language, IJAL 19.216–222 (1953).

TABLE 1

	Chm	Ya	Yu	Cho	Ji	Co	Wa	Mean % Cognates
Chimariko	—	38						
Yana	38	—	37	39	44	46	50	16
Yuma		37	—	35	35	48	48	19
Chontal		39	35	—	34	35	48	21
Jicaque		44	35	34	—	45	45	18
Comecrudo		46	48	35	45	—	55	15
Washo		50	48	48	45	55	—	13

40 C, there now are only six out of fifteen, plus the odd Chimariko-Yana. Seven are in the forties, and two in the fifties. This is as it should be, as compared with Uto-Aztecan, for this is a simple stock and Hokan a superstock.

The 35–39 C depths are confined to interrelations of languages situated on a NW–SE axis in the order: Chimariko, Yana, Yuma, Chontal, Jicaque; they include all interrelations on this axis yet computed except Yana-Jicaque (which rises to 44 C); but Chontal-Comecrudo is also 35 C. Evidently the linear arrangement of closer relationship ends with Chontal, whose 'onward' relations are about equal with Jicaque to the east and Comecrudo (Coahuiltecan) to the north—both of these incidentally being in Atlantic drainage. Comecrudo has no other low-depth coefficient, its other four running from 45 to 55.

One other Hokan language must be reckoned as off the NW-SE or 'Pacific' axis, as I am tempted to call it; namely Washo. This is a surprising fact because the edge of Washo territory, though over the Sierra Nevada in Great Basin inland drainage, lies only about sixty miles from Yana territory, and Washo was included in the original or Californian definition of Hokan. Nevertheless Yana and near-by Washo prove to be separated by 50 C, the greatest depth so far found for Yana! This is certainly surprising as against only 39 C for Yana-Chontal, with a separation of over 2000 miles. It is evident that the geographical proximity of Yana and Washo is a secondary and recent affair, and that they have been separated not only for a long period but probably by a considerable geographical interval much of that time.

Washo in fact proves to have been split off from every other computed Hokan language for 45 C or more. It is clearly in another division from the Pacific axis languages.

So on the whole is Coahuiltecan Comecrudo, though this has at least one relatively recent link, in the thirties of centuries, with Chontal.

However, Coahuiltecan and Washo, though both off the axis, have little in common with each other, since their time depth, 55 C, is the highest yet found.

We can then infer that the Hokan languages on which we have time depth figures fall into three provisional divisions: Pacific axis, Washo, Coahuiltecan, with the latter somewhat linked with the Mexican end of the axis, whereas Washo went off its own way, probably literally so, fairly early in the history of the superstock. That the lowest separation date for Washo in Nevada is with Jicaque in Honduras, 45 C, might contain a clue to its aberrant history, if it does not prove to be merely an accident due to the small numbers of stems—or languages—dealt with.

In the tabulation I have added for each language the mean of its cognate percentages. Coahuiltecan and Washo come out the lowest—which may be a reflection of their each being to date the only member of its division. The highest mean of cognate per-

centages is found in Chontal and next in Yuma. This may reflect the roughly central position of these two within the Pacific axis, even though they are separated by a long gap of 1400 miles.

I shall recur to intra-Hokan geography and history below, in order to compare them with Salish.

Of all the languages and stocks in the eastern United States that have been suggested as related to Hokan, Swadesh cites computations for only two, Atakapa and Chitimasha, and those ADJACENT on the Gulf: 45 C, 14%—much the same as for Yana and Washo. Presumably the relations of Atakapa and Chitimasha with Pacific Hokan will prove to be even more remote. Swadesh's provisional label for them as Eastern Hokanoid seems warranted.

5. While centuries running up into forties and fifties characterize the remoter relations within Hokan, all the relations of the (Macro-)Penutian languages as yet cited by Swadesh: Yokuts, Chinook, Tsimshian lie in the same range. The figures are: for Yokuts with each of the others, 55 C, 10%; but for these two northern idioms inter se, only 44 C, with 16%. It will be recalled that Yokuts is one of the five original Penutian families as brought together by Dixon and Kroeber in 1913, with a certain amount of supporting evidence in 1919, but that Chinook and Tsimshian were added later by Sapir, along with half a dozen Oregon-Washington 'families.' Of these, I long ago examined two of the geographically nearer, Takelma and Kus, and have re-examined them now, with even less noting of similarities to California Penutian in grammatical structure or vocabulary. True, this was only inspection; nevertheless it is clear that neither of these Oregon languages can be as close to any of the five Californian ones as these are to one another. As for Chinook and Tsimshian, the structure of both is so obviously different on the surface from what is common in the pattern of the Californian languages, that I have been inclined to attach little importance to occasional seeming similarities in vocabulary. In fact it was only respect for any linguistic judgment by Sapir that has kept me from dismissing the Penutian base of Chinook and Tsimshian as essentially speculative. Well, Swadesh is at last marshaling support for Sapir's perceptive intuitions—even though in this case it is as yet slender enough: around 10% in a brief basic vocabulary.

It is now also beginning to be clear what Swadesh's centuries of separation mean in general, in an overall approximate way.

CENTURIES	% BAS. VOC. COMMON	EXAMPLES	WHEN ACCEPTED
30	30	Uto-Aztecan	Long generally recognized
35	23	Pacific Hokan	First 'superstock' recognition
45	15	Other Hokan, E. Hokanoid, Chinook-Tsimshian	Later superstock recognition
55	10	Chin.-Cal. Penut.	Sapir's widest Penutian

On this basis, I would forecast that the five California Penutian stock-languages would prove, when tested, to come out as having separated around 35 centuries ago, or some, like Miwok and Costano, perhaps around 30.

Unfortunately, Swadesh has not yet worked out these more localized interrelations, with which I personally would have favored beginning, as being the most assured. But as his starting point is Sapir and not Kroeber, there is no just quarrel.

6. We come now to what Swadesh calls the 'Huave-Mayan phylum' (within which we have so far considered only the intra-Mayan Yucatec-Huastec relationship of 32 centuries). This group is comprised within what Mason and Johnson in 1940,[17] operat-

[17] J. Alden Mason, The Native Languages of Middle Am., 52–87; Fr. Johnson, The Ling. Map of Mex. and Centr. Am., 88–114; in C. L. Hay, *ed.*, The Maya and their Neighbors, 1940.

ing only for Mexico and Central America, classified as Totonac, 'Mi-zoc-uavean', Mayan, and Utaztecan, these four jointly constituting the southern part of Whorf's and their Macro-Penutian—'Megisto-Penutian' might have been an even more appropriate name. Here we really enter a vasty deep of ages. The lowest figure is 45 C, 14%, for Totonac-Zoque. From here, Swadesh's centuries run on up to 65, and his common basic percentages down to 6.4:

Yucatec-Zoque	55 C	10.4%
Yucatec-Huave	58 C	8.6%
Zoque-Huave	58 C	8.3%
Huave-Totonac	60 C	8.0%
Yucatec-Totonac	65 C	6.4%

There is an order here, and it is a reasonable geographic one: Totonac at one end, Maya at the other, Huave and Zoque between, but Zoque a bit the nearer of the two to Maya; and yet Huave and Zoque differ about as much from each other as from the two peripheral languages.

Nevertheless we are dealing with very low percentages of shared basic vocabulary; and the fact that the classification emerges as something reasonably possible should not make us forget that it also rests on precariously slender evidence. Six and eight per cent of a common basic vocabulary of 200 or fewer stems just is a very low number of discernible cognates. Part of it—six or seven words—might well be fortuitous resemblance, the rest—another six or seven—be due to borrowing, and nothing be left for the supposed genetic kinship. Deeper analytic penetration of the languages might easily reveal additional resemblances, but just as readily it might dispel as unsound some of those now counted as positive. Swadesh's skill and flair are indubitable, and he may well have steered his canoe of evaluation almost unerringly between the rocks of misjudgment. The value of his figures for these remoter relationships may lie precisely in their being suggestive of how proof will finally eventuate. But this is not proof; and I do not see how even the degree of statistical reliability can be expressed. If 15 instead of 12 Yucatec-Totonac cognates had been determined among 187 stems compared, the percentage would have gone up from 6.4 to 8.0 per cent, and the indicated time since divergence would have been five centuries less. Swadesh is well aware of this. When we come to these very low percentages of cognates, 1000 stems compared would give much solider assurance than 200; but with the added 800 we get increasingly outside of the vocabulary that is really basic and most stable, so that as the stems dealt with increase, the proportion of preserved cognates presumably goes down.

All in all, when we get definitely into the range of profound relationship that Swadesh, following Whorf and Trager, proposes to designate by 'phylum' instead of 'family' or 'stock', that is, from 5000 years up—as expressed by results of say 55 or 60 centuries or more—the quantity of evidence of lexical similarity is so tenuous that one wonders how much more credence to give it than to the Sapirian hunches.

This is not to say that co-membership in a phylum, and the age of a phylum to even more than 6000 years, may not be thoroughly demonstrable ultimately. What I am doubting is whether either is yet demonstrable from vocabularies of 200 items.

7. There remains the grouping for which Swadesh in his 1954 Anthropologist article cites the most numerous figures but also highly selective ones: Mosan—that is, Wakashan, Chemakuan, and Salishan. Within this group his results run up to 65 C, 7.1%, for Kwakiutl-Quileute. For Mosan the time distance is 86 C between the separation of reconstructed proto-Salishan and Quileute. I believe that Mosan has genetic reality, but primarily because of structural similarities, first noted by Boas and abundantly developed by Swadesh in a recent article.[18] It is when we come to the remote ages from which only small fragments sur-

[18] Mosan I: A Problem of Remote Common Origin, IJAL 19.26–44 (1953).

vive of the original stock of cognates, that structural similarities become our real resort. They scarcely amount to proof, but they may be sign posts. Thus, in Mosan, distributive plurals, feminine gender, numeral classifiers, abundance and variety of reduplication. Such patterns lack the precision of cognates based on phonemic equivalences, but they can nevertheless be fairly weighted for frequency, strength, or pervasiveness.

It is true that Boas cited such pattern similarities as possibly due to secondary influencing between original disparates, and therefore not allowable as evidence of genetic kinship. I would agree with Sapir and Swadesh[19] in regarding this alleged possibility as a technical rather than a likely one. Idioms and items of structure are what appear to be most open to invasion and acceptance between distinct languages; genuinely basic pattern or plan, least so.

In any event, if Boas is right, and structure, whether basic or itemized, can be imparted to languages of other types, the presumption is that enough vocabulary would be diffused with it to wreck many original cognates, and thus make lexical proof of original genetic unity increasingly difficult. In short, if the Boas position is accepted, it leads, as it was intended, to negativism in regard to our ever knowing much about the remoter history of all languages.

Allowing, however, for a more hopeful outcome, one other point can be made. If Swadesh's 86 centuries for Mosan prove to be approximately correct, what must we allow for the Mosan-Algonkin that Sapir thought of? At least 10,000 years, perhaps 12,000!

8. Within Mosan, Swadesh has published a previous article, in 1950, on Salish Internal Relationships,[20] which is much more meaningful than the small selection of results cited in his 1954 paper. In his 1950 table 2 he gives 325 values for the linguistic time distance between 26 Salish languages. It must be remembered that Salish was accepted by Powell in 1891, had been set up decades before, and has never been questioned or denied by anyone, including Boas, who in fact occupied himself intermittently for decades with a classification of the group. The remarkable thing about this 'conservative' stock is that its internal time distances run up by Swadesh's findings to above sixty centuries, or fully as much as the much less widely accepted Macro-Penutian or Hokanoid ones. The brief tabulation (Table 2) illustrates concisely this strikingly high age found for Salishan, presented from the angle of the number of separations at different age levels of the particular branch of Salish—South Georgian —which seems to remain least differentiated from original Salish.[21]

This tabulation of course is nothing more than Swadesh's 1950 table 2 high-lighted by being cross-sectioned. It agrees as fully with his table 3 and figure 1 (and presumably with his table 4 and figure 2 of reconstructed Salish), as does table 2. Swadesh's basic classification is into four divisions of Salish. Two of these are outliers to the north and south, both separated (though not far) from the rest of the stock: Bella Coola and Tillamook. The two other divisions form the main mass of Salish territory, which is continuous. They are Interior and 'Coast', the latter ramified into five branches, in order from north to south and then west: North Georgia, South Georgia, Puget Sound, Hood Canal (Twana), and Olympic.

What the tabulation shows is first of all the internal relationships of the recent Salish languages as seen from the point of view of that one of their eight groups which appears to be still the most similar to ancestral

[19] SJA 7.1-21 (1951), as cited.

[20] IJAL 16.157-167 (1950).

[21] The particular recent LANGUAGE which is most 'central' or least altered I infer, from 1950 table 2, to be: Lower Fraser.

TABLE 2

South Georgian linguistic distances (in centuries) from other Salish groups

Parallels in Other Phyla or Stocks	Centuries	Groups: Be Coo	N Geo	S Geo	Puget	Twana	Olym	Till	Int'r
		No. Langs							
		1	3	7	2	1	4	1	7
Eskimo	5–11			2					
	12–18			6					
? Ath	19–25			7	1				
Ut Azt	26–32		4	5	6				
Pac Hok	33–39		13	1	7	2			1
Macro-	40–46	3	4			5	5	1	5
Penut	47–53	3					20	4	28
Huave-	54–60	1					3	2	12
Mayan	61–67								3

reconstructed Salish. And second it shows with what internal relationships in non-Salish stocks and phyla these intra-Salish relationships most nearly correspond, as expressed in linguistic century-distances.

It appears that:

1. Languages within the South Georgian Coast Salish group may differ from one another about as much as the languages of the Athabascan family differ among themselves if these are really only about 20 C apart.

2. Adjacent South Georgian Coast Salish and Puget Coast Salish differ rather more than Uto-Aztecan languages differ among each other and perhaps somewhat less than Pacific (California-Mexican) Hokan.

3. South Georgian differ from North Georgian Coast Salish languages just about as much as Pacific Hokan ones inter se.

4. South Georgian differs from Twana Coast Salish, from the Bella Coola division, from Olympic Coast Salish, and from the Tillamook division about as the Macro-Penutian stocks differ from one another, progressively in the order mentioned; viz., South Georgia-Twana distance is about the same as Chinook-Tsimshian, South Georgia-Tillamook in the range of Yokuts-Chinook.

5. South Georgia Coast Salish differs from Interior Salish by a distance comparable to the larger Macro-Penutian distances and the Huave-Mayan ones—both these latter being reckoned as 'phyla' by Swadesh, whereas Coast and Interior Salish are only 'substocks.'

In short, it is abundantly evident that acceptance of Swadesh's method brings with it necessary acceptance of time depths within Salish definitely greater than in all other old-fashioned or Powellian stocks that have as yet been tested, and corresponding in range rather to the ranges within superstocks or phyla like enlarged Penutian, Hokanoid, and Huave-Mayan.

This situation is accentuated by the fact that while Boas regarded Sapirian Hokan, Penutian, etc. as so unprovable that he would not even seriously examine evidence on them,[22] he not only accepted the genetic unity of Salish but kept coming back to work on its classification—which in general was an uncongenial type of task to him.

9. What are we to make of this Salish anomaly?

One possibility is that there is something in the nature of Salish which brings it about

[22] He would not even examine the evidence presented for the much more modest California Hokan and California Penutian as originally set up by Dixon and Kroeber.

that when Salish languages have diverged for fifty or sixty centuries and even their basic vocabularies retain only ten or less per cent of cognates, there are non-lexical features remaining that enable us to recognize the kinship of Salish languages on inspection—that is, on early acquaintance, without detailed analysis and reconstruction. Such non-lexical features could only be phonological and structural patterns.

Another possibility is that the rate of lexical change is not uniform, but is more rapid in Salish than in most stocks. If for instance the basic vocabulary of Salish altered in as high a percentage of cognates in 30 centuries as that of Indo-European, Uto-Aztecan, Pacific Hokan, Mayan, etc., in 60 centuries, but the structural patterns altered only at the same rate in all of them, the Salish languages would then, after a seeming divergence of 6000 but actual divergence of 3000 years, be still alike enough grammatically to compel recognition of their kinship. The Pacific coast strong taboo on the name of the dead is readily invoked as a cause of more rapid lexical change—but perhaps too readily.

A third possibility is the influence of the fact that the 26 Salish languages occupy an almost unbroken and unusually compact territory of pretty heavy population. This situation might be a factor in two ways: intrinsically, in its effect on the languages themselves; or through its effects upon the Caucasians first coming into contact with them.

The intrinsic effect of a group of related languages continuing for long to occupy a compact block of territory is something we know very little about—it involves making comparisons of separate independent cases which are themselves comparative. Logically, however, it would seem that the effect would have to be conservative, both as regards structure and lexicon: change would tend to be slowed, especially for those languages near the center of the block. Consider, for instance: Lower Fraser is near the geographical center of the total Salish range and in contact only with other Salishan languages, and therefore sheltered from direct foreign exposures. Almost all influences reaching it from outside would necessarily be similar and cognate to those at work inside. Then imagine half the Lower Fraser population lifted up and dropped again in California, or in the Plains, and what would have happened to their speech after two thousand years of only non-Salish contacts. If preserved at all, this transported Fraser language would no doubt still be recognizable as Salish; but it would almost certainly differ from ancestral Fraser more than neighboring Squamish and Nootsak had meanwhile diverged.

The influence of alien contact on vocabulary appears very neatly in Swadesh's table 4, which lists the percentages which the several languages share with reconstructed Early Salish. I extract all the languages whose per cent is 35 or less; they are all on frontiers.

Bella Coola, 26: extreme N frontier, detached.
Tillamook, 31: extreme S frontier, detached.
Coeur d'Alene, Columbia, Kalispel (Flathead), 32, 33, 35: interior frontier toward SE
Lower Chehalis Quinault, 32, 33: farthest W; Chinook, Quileute, Athabascan neighbors.
Pentlatch, 31: on Vancouver Id; Nootka neighbors[23]

The effect on explorers and travelers of encountering related languages, gradually changing more as the mass of their territory is penetrated farther, can be illustrated from Swadesh's Salish table 1. Suppose first contact had been established with the Coeur d'Alene, and progress were northwestward toward the coast. The Coeur d'Alene share 55% of their basic vocabulary with Kalispel and Pend d'Oreille, 44%

[23] Swadesh's fig. 2 suggests that there has been a geographical transposition of Pentlatch and Comox (37%), which latter has in the historic period been on the Salish frontier and in contact with Kwakiutl and Athabascan as well as with Nootka, whereas formerly it seems to have been Pentlatch that was peripheral.

with Okanagon. Okanagon in turn shares 47% with Thompson, Thompson 57% with Lillooet, which in turn retains 23% of basic Coeur d'Alene. With Lillooet we leave the Interior division and enter the Coast. Highest here are the nearest: Fraser 28%, Nootsak 26% shared with Lillooet—higher than the 23% Lillooet-Coeur d'Alene figure within the Interior division. Nisqually 24%, Twana 23% follows, and no language in the North or South Georgia or Puget branches is less than 20%. The Olympic branch runs 21–17%, the distant Bella Coola and Tillamook divisions 19% common with Lillooet. Or, if we shift from Lillooet as a base to Fraser as its closest cousin on the Coast, then Bella Coola and Tillamook are up to 23%, Twana to 31, Puget 33–36, North Georgia 35–42, South Georgia 36–75%. It is evident how continuous the connecting threads are, and how any one acquiring even a consistent smattering of the Salish languages could not help recognizing that they all contained a common element.

It is also evident that the Salish situation is generically similar to that in Algonkin, except for the Algonkin population being more than three times and the Algonkin area perhaps ten times larger than the Salish. And again the most divergent languages are peripheral: Arapaho, Blackfoot, probably Cheyenne, Beothuk if it belongs at all.

Similar also in a general way, but again of a much wider range spatially, is the Uto-Aztecan picture; except that here Athabascans and Yumans separate—by a not too wide gap—the 'Shoshoneans' of the United States from the Piman, Taracahitian, Nahua, and other essentially Mexican languages of Uto-Aztecan lineage.

10. Hokan however presents a different picture. I am confining myself to 'Pacific Hokan', which was the first to be recognized, whose geographical distribution is essentially linear, and in which the few linguistic depths that Swadesh has computed nearly all run between 35 and 40 centuries. In sequence from north to south we find the following, all originally reckoned as of separate stock.

In northernmost California, five stocks: Chimariko, Karok, Shasta, Achomawi (these two were united by Dixon), Yana. There were two or three aberrant local forms of Shasta on which too little is preserved to judge whether they represent languages or poorly selected and recorded dialects. Within Achomawi, Atsugewi is a well separated language. Otherwise, all variants with each of these stocks are within limits of intelligibility. In short, the stocks are contiguous in a three-quarter circle; the distance between Chimariko and Yana across Penutian territory is something over 50 miles; that between the western boundary of Karok and the eastern of Achomawi, well under 200 miles, airline.

(West of Yana, separated by 60 miles of Penutian Maidu, begins Washo territory, but I am excluding Washo from Pacific Hokan.)

South of Chimariko about 85 miles begin the Pomo, with Athabascans and Yuki between.

South of the Pomo about 140 miles, and separated by Penutians and San Francisco Bay, begins a block of three coastal Hokan stocks, Esselen, Salinan, and Chumash.

From the most southeasterly Chumash to the nearest Yuman language, Diegueño, it is 100 miles, the southern Californian Shoshonean languages intervening. The Yuman block, in California, Arizona, and Lower California, is the largest territory in Pacific Hokan.

Some 150 or 160 miles south of the Yuman Cocopa in Sonora begin the Seri. Other Yumans are nearer the Seri, across the Gulf of California, in the cul-de-sac of Baja California.

Next comes the longest gap or interval, just about 1000 miles from the Seri to the Tlapanec, on the Pacific Coast in eastern Guerrero.

The next jump, Tlapanec to Chontal or Tequistlatec, is about the same as Yuman to Seri, around 160 miles. The Chontal are on the Oaxaca coast, somewhat short of Tehuantepec.

Another interval is from Chontal to Jicaque on the Atlantic side of Honduras, about 500 miles.

The last gap brings us to the Subtiaba in western Nicaragua, about 150 miles south of the Jicaque. Subtiaba appears to be similar to Tlapanec, as if the two had been dispersed rather recently.

It is evident that this Pacific Hokan distribution is the very antithesis of Salish distribution. Instead of one roughly quadrilateral block with two small and near-by outliers, we have nine separate blocks of Pacific Hokan territory, none of them large and only the Yuman one medium-sized. They stretch over 3000 miles, with intervening gaps of 50, 100, 150, 500, and 1000 miles. Moreover, the nine areas contain fifteen groups reckoned as separate families in Powell's time, as against the then admitted familial unity of all Salish languages. Yet the available computations for Pacific Hokan run under 4000 years (with the one exception of far-distant Yana-Jicaque at 4400), the internal Salish ones up to 6000!

If the dozen or more former families or stocks now included in Pacific Hokan continue to run like their first samples, with a separation time mostly of 35 to 40 centuries, the history of the societies of this Hokan division as a whole must have been very different from that of the Salish. These latter apparently have always formed a solid geographical mass, as far back as we can infer anything about them, without migrations or larger movements, and probably about where they were found by the first Caucasians. There were internal shifts of tribes within the total Salish aggregate, no doubt, and gradual or intermittent expansions or contractions in the frontier of this total aggregate. But the modern interrelations are so continuous and so conformable to position on the map today, that the differentiations must have been continuous and steady also, on the whole, without many violent geographic displacements. Dialects seemingly diverged into languages, languages into branches, branches into grand divisions. What was originally essentially one Salish language differentiated into something of family scope, this into a stock of several families, in Swadesh's suggested terminology, and the stock into a subphylum, whose 'divisions', if it were not for their contiguity and intergradations, might elsewhere have been first classified into separate stocks.

Pacific Hokan, contrariwise, at one time evidently suffered an explosive disruption. It had probably already branched into ten or a dozen languages—there may have been more, some having become extinct since. Then these became dispersed over nearly three thousand miles, into what finally ended up as nine separate areas. The dozen or more languages were at that time roughly about equally different from one another, to judge from their appearance and from the few time-depth computations available. They became scattered far enough that most of them were no longer able to influence and hold one another steady. The main scattering may have happened fairly late, since most of the dozen stock-languages have notably few internal variants of more than dialect strength. Or they may have been on the defensive and have lost territory, so that dialects were obliterated or merged. Except for Yuman and Pomo, the Pacific Hokan languages look as if they were shrunken, compacted residues of once larger speech groups. Hence their monadal quality, without intergradations or salient indications of grouping into larger classes.

The five groups in Pacific Mexico-Honduras-Nicaragua might conceivably be the result of one major migration, of groups which, though already differentiated, shared common fortunes, but dropped off the route one by one on the way. Or will they prove to have enough in common to suggest that something like 3000 or 3500 years ago they

were still essentially one, but, losing all contact with one another, they became as different as most of them now seem to be?

Ten of the fifteen Pacific units remained wholly or partly in California. It is a more economical supposition that most of them were there originally than that they all wandered in or were driven together into California. Their particular distribution confirms this. Five tiny stocks surround the northern head of the great interior valley as if the Penutian Wintun had crowded them back into the hill country. It seems hardly possible that five such distinct yet related stocks could have originated in such close contact and continued to diverge as much as they have, and without intergradations. On the two sides of the same northern arm of the valley are the Pomo in the Coast Ranges and the Washo across the Sierra Nevada, though these last seem not to be specifically of the Pacific Hokan division. Then, to the south, Penutians have reached the coast in addition to holding the interior, until, beyond Monterey, the three-stock block of Esselen-Salinan-Chumash coastal Hokans begins. These in turn are separated from the Yuman block by what looks like a Uto-Aztecan expansion around Los Angeles —the only stretch of Pacific coast in the United States held by Shoshoneans in historic time.

These are specific historical reconstructive adumbrations. They do show what vistas are opened up as soon as linguistic distributions on the map are added to linguistic time depths. They also show that time computations must always be read with reference to the map, not in vacuo. The much greater time depths of the Salish than of the Hokan languages would be misconstruable in the abstract, as indicative of less proneness to immanent or spontaneous change, in view of the fact that Salish languages ever since first known have been accepted as a genetic unit whereas the Hokan ones were long not recognized as affiliated with one another at all. In fact the time depth findings might be misleading if they had not been brought into relation with Salish intergradation and territorial continuity and Hokan linguistic and geographic discontinuity.

My question above as to why the difference between the histories of the two stocks, I shall leave as a rhetorical one—that is mainly unanswered: it is clear that fuller precise knowledge can hope to answer them. Some of the principles or general considerations involved are evident now; whereas present pushing further of particular reconstructions would be premature.

11. At one point I have consciously simplified the foregoing problem: in restricting the discussion to Pacific Hokan, when there is also not only Gulf Coast Hokan (Comecrudo, Atakapa, Chitimasha) but also possible Hokan-Siouan beyond (Siouan, Muskogian, Caddoan, Iroquoian). I feel that solid progress from a new method like glottochronology will best eventuate by our first applying it mainly to bodies of knowledge which have already been organized fairly well, to genetic groups that are sure, and then proceeding from these with further comparisons. The glottochronological study that will be of most enduring utility of those so far made by Swadesh is probably the Salish one: it is a terra firme to which further terrain can be added. Considerations of the remoter relations—except for an occasional flyer to satisfy personal curiosity—are best deferred. The relations of the ten or twelve principal branches of Uto-Aztecan are far more important now than the relation of Tanoan to Uto-Aztecan because they refer to a considerable systematized body of knowledge to which Tanoan is at present still mainly tangential. And if Zuni is really related to Aztec-Tanoan, it is so remotely related that our understanding of the nature of its relation will long be so shadowy as to be of more speculative interest than significant import.

Similarly as regards Yuki. I have myself worked on this puzzling little stock, but feel that knowing whether it is primarily Hokan

or primarily Penutian, or basically the one but secondarily influenced by the other, is not really going to be an important answer, because the relationship will in any event be ancient, thin, and vaguish. A knowledge of the time depths of Pacific Hokan, or of the solid block of California Penutian, will help solve many more other problems than anything we are likely to determine about a small erratic like Yuki. Its placing is an intriguing challenge to our virtuosity; but Penutian and Hokan, Uto-Aztecan and Mayan, Algonkin and Athabascan, represent definite bodies of systematizable understanding of large histories.

As an exercise in imagination and interest, an occasional survey by archaeologists and ethnologists of what conjectural dates of theirs seem to match with glottochronological ones, may be provocative. We seem still to be far from the point where such matching can lead to solid conclusions. Suppose that two languages now spoken in California were just budding away from each other 5000 years ago, and that the archaeologist has evidence that the replacement of the metate by the mortar in California also was happening then. The linguistic divergence may actually have happened elsewhere than where the shift in grinding technique occurred, and both languages have moved in later. And there almost certainly were a number of other languages spoken in California, whose associated culture also underwent the metate-mortar shift. If two separate techniques of dating different kinds of historical data both emerge with a specific result like 5000 years ago, even with some vagueness as to whether the two events happened in the same area, an illusory sense is stimulated that the two dates are significantly connected, when the probability is much greater that there is little or no real relation between them. How can an archaeologist tell whether the people who left a certain run of artifacts talked Salish, Shoshonean, or Hokan, or what branch of Hokan? Not that the possibility of such association being proved is utterly hopeless. But we cannot begin with a few haphazard events connected merely by having an age of so and so many millennia in the same general area. The chances are much better of the association being real if both dates are only 500 years old. After that we may be able to proceed to 1000, and perhaps finally to 5000 or even 10,000 years. Progress of knowledge that stands up will be step by step. It is to be hoped that Swadesh will be able to resist the importunities and pressures to which he will be subjected by anthropologists who are nonlinguists but ready and eager to use linguistic results.

What I hope here to have shown the need of is that primary concern be with relatable sets of phenomena: of vocabulary with structure, and of languages with their known historic territories; of a set of datings with a history of linguistic developments. The time depth of a group of languages that are obviously related, and have long been recognized as such, can presumably now be worked out with reasonable precision, and will normally be far more significant than a scattering of long dates for separations of languages whose very connection is novel and startling but tenuous. When we shall have half a dozen stocks known with the fullness of coverage that Swadesh has applied to Salish, comparison of these half dozen will give us not only a large piece of coherent history of native American speech but will help to refine the glottochronological method and probably to give us unsuspected new principles and general insight. AFTER that let us call in the archaeologists, the geneticists, and whoever is interested, and see what we can forge jointly.

We appear to be nearing the point at which a planned, systematized, and cooperative attack on the glottochronology of one continent, presumably for tactical reasons North America, is in order. It is to be hoped that the coordination of this enterprise be done under the leadership of the scholar whose imagination and drive originated the method and have carried it so far.

LINGUISTIC AND ARCHEOLOGICAL TIME DEPTH IN THE WEST INDIES

DOUGLAS TAYLOR
DOMINICA, B.W.I.
AND
IRVING ROUSE
YALE UNIVERSITY

0. Introduction
1. Time depth of Island Carib
2. Archeological correspondences

0. In a recent symposium on "Time Depth of American Linguistic Groupings," Swadesh[1] published a date for the separation of the Dominican and Lokono languages of the Arawakan stock based on a word count by Taylor. The following is a revision of this date by Taylor (1) and a consideration of its archeological significance by Rouse (2), together with some discussion of the question of migrations into the Antilles, which developed as a by-product of the dating problem.

1. It is here assumed—but can, I am convinced, be proved—that the language of the historical, so-called Island Carib is Arawakan, and almost as closely related to Lokono or True Arawak of the present time as is Roumanian to French.[2] It is further assumed that this language is a continuation of the one spoken by the 'Igneri' or pre-Carib-conquest inhabitants of the Lesser Antilles, whose womenfolk were spared, according to seventeenth-century tradition and in all likelihood, to become wives of the conquerors and mothers of subsequent generations of 'Island Caribs.'[3] French missionaries writing in the middle of the seventeenth century tell us that the same language was spoken in all the islands then inhabited by this people; from which we may at least conclude that dialectal differences were not so great as to preclude mutual intelligibility. At the end of the eighteenth century this language was taken to Central America by some five thousand 'Black Carib' deportees; and is spoken there today by about six times that number of individuals.[4] In the Lesser Antilles it lingered on only in St. Vincent and Dominica, where it became extinct about 1920. Apart from the modern Central American dialect, only two records of this language are known to be extant: one, fairly full, dating from the middle of the seventeenth century;[5] the other, very sketchy, from the end of the nineteenth century; both made in Dominica.[6]

Phonetically, the Dominican dialect – even in its most recent stage—was much closer to modern Lokono than is that spoken by the present-day Black Caribs; and for that reason I at first employed it when seeking to estimate by means of lexical statistics the length of time that must have elapsed since Lokono and Island Carib were one and the same language. However, there are

[1] Morris Swadesh, Time Depth of American Linguistic Groupings, AA 56.361–4 (1954).

[2] C. H. de Goeje, Nouvel examen des langues des Antilles . . ., JSAP 31.1–20 (1939); reviewed by Taylor in IJAL 17.257–9 (1951).

[3] Douglas Taylor, Diachronic Note on the Carib Contribution to Island Carib, IJAL 20.28–33 (1954).

[4] Douglas Taylor, The Black Carib of British Honduras, VFPA 17 (1951).

[5] Raymond Breton, Dictionaire Caraibe-François . . ., édition facsimile Jules Platzmann (Leipzig, 1892); Dictionaire François-Caraibe . . ., édition facsimile Jules Platzmann (Leipzig, 1900); and Grammaire Caraibe suivie du catéchisme Caraibe, nouvelle édition, L. Adam et Ch. Leclerc (Paris, 1878).

[6] Joseph Numa Rat, The Carib Language as now spoken in Dominica, West Indies, JRAI 27.293–315 (1898).

several disadvantages in using an extinct language for such a purpose; one cannot possibly obtain a word missing from the record, or decide which of two apparent synonyms was the commoner. So, for example, confronted with the item *fall* of Swadesh's basic word-list, I could find nothing in the record of recent Dominican to match with Lokono atikida; while Breton's seventeenth-century dictionaries offered an embarrassing wealth of different equivalents employed in different contexts, among which two seemed to be equally 'basic': áikua and atíkera. So that, although the latter is obviously cognate with the Lokono word, this item had to be dropped from the list because there was no indication that it was the commoner of the two, or differed from the former in meaning, as for example *tumble* differs from *drop*. Since the living dialect offers only one word, áigua, this item has been included in the list, and scores a MINUS.

Nevertheless, instead of a cognate ratio of 62/146, as reported by Swadesh,[7] I now get a ratio of 72/154, or 46.75%, which indicates a time depth of 18 (instead of 21) centuries. The English words whose Lokono and Island Carib equivalents are matched all appear in one or another of Swadesh's test-lists; but as these call for 200 items, and as Lokono data are still not abundant, I have included some items that do not appear in all of them. Even so, I can match only 186 pairs; and for greater precaution 32 of these have been queried and omitted from the count. Such precaution seems advisable wherever the evidence for cognation remains inconclusive one way or another, and wherever there is uncertainty as to usage. So, for example, apart from the pair *vomit* ereda : euéra, I can find no support for the existence of an r : u correspondence, though there is plenty of proof that d : r is regular; while in such a case as that of the words for *cut*, Island Carib (and probably Lokono) has a number of different terms whose employment depends on whether the cutting is done with a knife, scissors, an ax, etc.

As Swadesh[8] has been careful to point out, the results that have already been obtained from lexico-statistical dating, although distinctly encouraging, must still be regarded as provisional in greater or lesser degree. So far as languages with little or no written history are concerned, there appears to be a real danger of over-estimating time-depths through failure to recognize actual cognates as such. For it cannot be supposed that the Indo-European family of languages is unique in containing such divergent products of phonetic change as (Russian : English) jazík : tongue, or (English : French) sew : coudre, blow : souffler, or (English : Spanish) eat : comer. Yet the student of an American Indian linguistic family rarely has any knowledge of such 'links' as Latin consuere, connecting *sew* with *coudre*; Latin com-edere, connecting *eat* with *comer*. Lokono : Dominican Island Carib *my father* dati : núkusili look even less like cognates than when núguci of the modern Central American dialect is substituted for the latter word; yet t : s for the former and t : c for the latter dialect are attested in some other cases, such as *dig* atika : asíka : acíga; while the former correspondence suggests borrowing from related Goajiro, in which s regularly corresponds to Lokono t. According to Hickerson, Lokono has a morpheme -ku- *consanguineal elder*; and the Island Carib forms meaning *father*, *mother*, and *grandfather* almost certainly contain this morpheme in a fossilized state. Similarly, the pair *fly* amoroda : áhamara are seen to be cognate only when compared to amą́hara *fly* of the Dominican dialect, both the latter having hamą- as stem alternant.

To judge by the records of Daniel Brin-

[7] Swadesh, *op. cit.*, p. 362.

[8] Morris Swadesh, Archeological and Linguistic Chronology of Indo-European Groups, AA 55.349–52 (1953).

ton,[9] de Goeje,[10] and Hickerson,[11] Lokono (True Arawak) exhibits a rather high degree of phonetic instability, so far as this comparatively short and recent period is concerned. So, for *two*, the first and second of these authors give biama, but the third has bian ~ biani-; for *three* the first gives kabuhin, the second kabuin, and the third kabun ~ kabuni-; for *four* the first has both bibuti and bibiti, the second bibiti alone, and the third bibiti ~ biti ~ bis-. The personal pronouns are given by Brinton as dakia ~ dai *I*, bokia ~ bui *thou*, likia *he*, turreha *she*, uakia ~ uai *we*, hukia ~ hui *ye*, nakia ~ nai *they*; and by Hickerson as dai, bui, lira, tora, uaiko, hui, naira. Brinton's -piru corresponds to -firo, meaning *big* or *strong*, of both de Goeje and Hickerson; the word for *blood* is given by de Goeje as ite, but by Hickerson as utu. Brinton tells us that 'from kasan *to be pregnant*' (cf. -sa ~ osa *child*) 'comes kasaku *the firmament*, big with all things which are, and kasahu behu *the house of the firmament, the sky, the day*.' This last phrase renders highly plausible a relationship between modern Lokono kasakabo *day* and Island Carib ubéku > ubéhu > ubéu *sky or clouds* (the first form attested from early Dominican, and the last two from the Central American dialect of, respectively, one hundred years ago and today). Nor has Island Carib always been much more conservative, to judge by the (early Dominican : modern Central American) correspondence, binálepule : bináfi *morning*; although here a lapse of three hundred years and a change of habitat are involved.

Evidence for the cognation of some of the rejected pairs, though inconclusive, is not lacking. So, Taino da-guita *my rope* (in which the g is probably a dummy) suggests that the final -ho and -ru of (Lokono : Island Carib) taho : fitaru *rope* were once suffixes, and that these words are cognate. The resemblance of the Island Carib forms meaning *father* (-guci) and *mother* (-gucu) to one another makes it likely that Lokono -io *mother* resulted from the weakening of older *-tio, more closely paralleling Lokono -ti *father*; and that if so, also Lokono : Island Carib ioio : dódo- *wet* are cognate, and come from older *tiotio. If I had to reduce the number of queried pairs omitted from the count, I should score a PLUS against those meaning *hold, left, man, mother, old, rain, root, rope, sky, snake, suck, vomit, wet, worm*; a MINUS against those meaning *cut, fruit, fur, guts, knee, near, short, thin, walk, when*; and leave the query against those meaning *earth, hot, hunt, I, if, in, not, what*. This would raise the cognate ratio to 86/178, or 48.3%, which indicates a reduction of some 75 years in the time depth. It seems to me unlikely that further data will reveal a percentage of cognates much higher than these last figures; but even in the improbable event of all the doubtful pairs proving to be cognate, we would still have to reckon that these two languages had been diverging for the past 13 centuries.

By substituting the early Dominican for the modern Central American dialect, a cognate ratio of 79/154 is obtained; seven more agreements are registered (*bird* ulíbịu, *five* uákabu ápurku, *four* biąburi, *grass* kálau, *husband* iráiti, *alive* kako-, and *person* iláku), the items meaning *guts* and *hunt* each score a MINUS, and those meaning *dust* and *fall* must be queried. When the difference in dates is taken into account, this reduces the time depth found for the modern dialect by less than half a century. By using recent Dominican in the same way, three more agreements are registered (*four* biąburi, *grass* kárau, *husband* iráiti), but the number of pairs matched is considerably reduced through lack of data, which probably accounts for the time depth of 21 centuries as reported by Swadesh. It is perhaps superfluous to add that the absence of

[9] Daniel Brinton, The Arawack Language of Guiana in its Linguistic and Ethnological Relations, APS-T, 14.427–44 (1871).

[10] C. H. de Goeje, The Arawak Language of Guiana (Amsterdam, 1928).

[11] Nancy P. Hickerson, Ethnolinguistic Notes from Lexicons of Lokono (Arawak), IJAL 19.181–90 (1953). Also in personal communication.

cognate pairs of words with the same meaning does not necessarily imply the absence of cognates in the two languages; so, for example, Lokono kudibiu *bird* and C. A. Island Carib guríbiua *sp. of bird* (unidentified), Lokono ibitua *burn* and Island Carib ébeda *kindle*, Lokono iribe- *dirty* and Island Carib iríbi *soot*, Lokono buin *whole* and Island Carib buį *full*.

One apparently unrelated pair of words, (Lokono : Island Carib) dakia ~ dai : nugúia *I*, is of particular interest for the prehistory of the Antilles. Mason tells us that in Arawakan languages 'the first person pronoun is usually *nu*, whence the generic name *Nu-Arawak*'; whereas Lokono has dakia ~ dai (and affixial dA- and -da, *vs.* Island Carib nV- and -na), Goajiro has taya (and tA-), and Taino has variously spelt dacha, daça, daca (and dA-). These latter forms (those beginning with an apical stop) would therefore appear to be innovating as compared with the former (beginning with n), among which are those of far-away Campa (nu) and Amesha (na˙). Moreover, it is difficult to see how either Taino or Igneri and Island Carib forms meaning *I*, *me*, *my* could have been borrowed after migration to the Antilles (as speakers of English borrowed *they*, *them*, *their* from Scandinavian), since the Taino were separated from other da-speakers by the Igneri or Island Carib, and the latter from other nu-speakers by the Lokono. And if, as seems likely, they were inherited, it is reasonable to assume for the linguistic forbears of Lokono, Goajiro, and Taino a period of common development in which those of the Igneri and Island Carib had no part. Compare:

	Taino	Lokono
I	dacha, dA-	dakia, dA-
we	guA-	uakia, uA-
he	nV-	likia, lV-
	Goajiro	**Island Carib**
I	taya, tA-	nugúia, nV-
we	waya, wA-	uagía, uA-
he	nia, nV-	ligía, lV-

Should this view be correct, it would follow that the separation of Igneri and Lokono preceded that of Taino and Lokono; so that Igneri and Taino could not then possibly be recently diversified forms of one and the same parent tongue. Unfortunately, the extant Taino vocabulary is too meagre, the transcription too erratic, and the given meanings too uncertain to offer much help in this respect. Nevertheless, comparison of the following Taino items (with, so far as possible, the spelling found in the sources) with their Lokono and Island Carib equivalents shows more agreements with Lokono (and with Goajiro) than with Island Carib of either dialect or period. So, the forms meaning *I*, *know*, *man*, *stone*, *three* agree only with Lokono, those meaning *dog* and *woods* only with Island Carib, while that meaning *four* agrees with Lokono and with the Dominican dialect of Island Carib only. *black* xey-, *breast* too, *dog* aon (also : alco), *ear* -rique-, *eye* -aco, *good* tayno, *I* dacha (or daça or daca), *know* ita, *man* guaxeri (*cf.* Goajiro hašič˙i *man*), *one* hequeti, *say* ahia-, *see* oca-, *skin* -ra, *stone* cibe, *tooth* -ahi-, *two* yamoca, *four* yamoncobre, *rope* -guita, *sea* bagua, *sky* turey, *three* canocum, *woods* arcobuco. The name for the Tainos' ceremonies, areito, which those who witnessed them described as 'bailar cantando,' may be compared with Goajiro airøha *to sing*. Taino higuera *calabash* (*Crescentia cujete*), with which compare Lokono iuida, Island Carib uíra (with the same meaning), probably contains the Taino word for *fruit* (*cf.* Lokono iui *fruit*) in composition with the stem meaning *skin* (-ra in Taino, -da in Lokono), or with that meaning *tree or wood* (also -da in Lokono, but unattested for Taino); although the Island Carib word was not analyzable even in the earliest recorded dialect.

This seems to imply that the Antilles were peopled by two distinct migrations of different Arawakan tribes. If the Igneri came first, we may presume that this

people left the mainland some time during the first three centuries of this era; to be followed about 700 years later by the Taino. In this case, it seems unnecessary to assume that any 'conquest' or fighting took place at that time; for in 1635 the Island Carib of St. Vincent allowed two shiploads of escaped African slaves to settle in a part of their country, where they long continued to be joined by fugitives from neighboring islands; and it is by the descendants of these Negroes that the only surviving dialect of Island Carib is spoken today. But the Taino, as new comers, would have been expected to 'move on' if and when increase of population rendered means of subsistence scarce. Thus they would have remained a minority, and have been gradually assimilated by the Igneri in the smaller, more southerly islands; while in such a comparatively large island as Porto Rico they would have come to outnumber and assimilate the few Igneri that may have preceded them thither before any pressure of population was felt.

If, on the other hand, the Taino were the first to leave the mainland, I doubt (on account of the apparent time-depths between the languages concerned) whether this could have been much earlier than A.D. 600, or much later than A.D. 1000. And in this case, the Igneri must have followed rather closely on their heels, and chased them into the Greater Antilles; for geographic distribution of the languages then compels us to assume a 'conquest' of the Lesser Antilles by the Igneri, since it is unlikely that the Taino should have allowed themselves to be 'assimilated' or have given up their old homes to the new comers without a struggle. Moreover, this second, alternative solution raises yet another problem,—that concerning the separation, long before any Arawakan migration to the islands, between Igneri and the common progenitor of Lokono, Goajiro, and Taino.

In general, Hickerson's Lokono forms have been preferred; so (Lokono : Island Carib) *mouth* -leroko : -iúma scores a MINUS, although de Goeje lists obviously cognate uima ~ uma ~ ema with the same meaning. That the former, and not the latter, is the ordinary Lokono term is clear from two recordings cited by Brinton, the one (dalerocke *my mouth*) dated 1598, and the other (daliroko *my mouth*) dated 1800. In a few instances, however, Hickerson's form has been rejected, as in the case of her -doli *root*. De Goeje also lists this last form, but gives it the special meaning of 'racine tuberculeuse' (which Island Carib -ílagola does not have), as opposed to ekura *root, vein, nerve*, which tallies semantically with the Island Carib. This is one of the items which I have thought better to omit from the count.

In the following word list, the Lokono forms follow the English, and are followed by those of the C.A. dialect of Island Carib.

all -makua : sú(gubai) (−), *and* kena : -úma (−), *animal* iliki-n : ilógo-ni (+), *arm* -duna : aróna (+), *ash* balisi : balígi (+), *back* -iabo : -anágą (−), *bad* uakaia- : uríba- (−), *bark* (ada) uda : t-úra (uéue) (+), *because* -doma : -rúma (+), *belly* -dibeio : -urágai (−), *big* fi-ro : uáiri (−), *bird* kudibiu : dunúru (−), *bite* akota : ágora (+), *black* karime- : urí- (−), *blood* utu : h-íta-o (+), *blow* afudi- : á-fu-ra (+), *bone* -bona-ho : -ábu (+), *breast* -odio : -úri (+), *breathe* ahakobu- : auára-gua (−), *burn* ibitua : á-guda (−), *child* ilon-/iren- : iráho (+), *claw* -obada : -úbara (+), *cloud* ororo : ubéu (−), *cold* meme : dili- (−), *come* anda : (n)iób(u)i (−).

cut aroka / asoka : íbiha / ácuga (?), *dark* bura- : buri- (+), *day* kasakabo : uéiu (−), *die* ahoda : a-hila-ra (+), *dig* atika : a-cíga (+), *dirty* iribe- : uíe- (−), *dog* pero : áụli (−), *down* onabu / abomun : ónabu / -ábugię (+), *drink* ata : áta (+), *dry* uato : mábai- (−), *dụll (not sharp)* ma-mana- : ma-mána- (+), *dust* akumuiu : t-ífe múa (−), *ear* -diki : arígai (+), *earth* uaia : múa (?), *eat* eke : áiga (+), *egg* karinasa : gáię (−), *eye* akusi : águ (+), *fall* atikida : áigua (−), *far* taha : díse (−), *father* -ti :

-guci (+), *fear* hamaroka- : -anúfude (−), *feather* -bara : t-ubána (dunúru) (−), *few* kabun : m-íbe (−), *fight* afara : a-gá-ią-dagua (−), *finger* akabo ibira : l-iráo -úhabu (+).

fire hikihi : uátu (−), *fish* hime : údurao (−), *five* aba dakabo : ségo (−), *flesh* isiroko : ógorogo (+), *flower* koro-ho : íleue (−), *fly* amoroda : áhamara (+), *foot* -koti : ugúdi (+), *four* bibuti, bibiti ∼ biti ∼ bis- : gádoro (−), *fruit* iui : h-į (?), *fur* -iti : h-íu (?), *give* isika : íciga (+), *good* osa : buídu (−), *grass* karau : sagádi (−), *grease* kihi : agólei (+), *green* imoro- : urígi- (−), *guts* -ite : -isása (?), *hair* -bara : ídiburi (−), *hand* okabo : úhabu (+), *he* likia / lira : ligía (+), *head* isi : icígo ∼ icógo (+), *hear* akanaba : agąba (+), *heart* oasini : anígi (−), *heavy* kudu- : hóro- (+), *here* ia-ha : iá(-ha) (+), *hit* aboroka : áfara (−).

hold abokota : ágoda (?), *hot* tere : hára (?), *how* halika- : ída -ią (−), *hunt* aioka : áibaha, a-gá-liru-ha (?), *husband* ireti : úmari (−), *I* dakia ∼ dai : nugúia (−), *if* -faroka : áhą- (?), *in* oloko : -ídą / íla-o (?), *kill* afara : áfara (+), *knee* -kuru : -gácorogo (?), *know* aita / adita : a-súbudi-ra (−), *leaf* -bana : -ubána (+), *left* -baro : -ubáuna (?), *leg* -dana : -úruna (+), *lip* -leroko : -iumáru (−), *alive* kake- : n-ibága-i (−), *liver* -bana : -ubána (+), *long* uadik- : migífe- (−), *louse* ie : ję-į (+), *man* uadili : eiéri (?), *many* ioho : g-íbe (−), *moon* kati : háti (+), *mother* -io : -gucu (?), *mountain* fororo : uóbu (−), *mouth* -leroko / -noroko : -iúma (−).

name iri : íri (+), *near* -erebu : -rúgabu (?), *neck* -noro : -igína (−), *night* oriko : áriabu (+), *nose* -siri : ígiri (+), *not* ma- / koro : ma- / máma (?), *old* uahadu- / hebe- : uái-ha- (?), *one* aba : ábą ∼ ábana (+), *person* loko : mútu (−), *play* ibiraka : hurára- (−), *pull* aduruda : a-cáua-ra (−), *pus* oko : óho (+), *rain* oni-kia : húia (áhuia) (?), *red* kore- : funá- (−), *right-hand* -isa : -áuere (−), *ripe* kore- : funá- (−), *river* onikain : dúna (−), *road* abona-ha : óma ∼ -éme-ri (−), *root* ekura / doli : ílagolá-o (?), *rope* taho : fítaru (?), *rotten* toro / kolo : niąla- (−). *saliva* uraroni : aróre-i (+), *salt* pamo : sálu (−), *sand* motoko : ságaų (−), *say* adia-: erégа (∼ arįaga) (+).

sea bara : barána / baráua (+), *see* adoka : aríha (+), *seed* -si : -į / -íla (−), *sew* akosa : áhoca-gua (+), *sharp* kamana- : ga-mána- (+), *short* baseken- / auka- : dú- / láugua- (?), *sit* abalta : niuru- ∼ iú (−), *skin* uda : úra-o (+), *sky* kasa-hu behu : ubéu / siélu (?), *sleep* adonka : arúmuga (+), *small* sioko : nió-rao- (−), *smell* adimisa : irímica (+), *snake* uri : óui / héue (?), *some* aba-no : -íbiri (−), *stand* adinama : rárama- (−), *star* uiua : uarúguma (−), *stick* (*N*) ada : uéue (−), *stone* siba : dóbu (−), *straight* misi- : surú- (−), *suck* atuku- / asoroto- : su- / sura- (?), *sun* hadali : uéiu (−), *tail* ihi : íli (+), *that* tora / lira : túra / líra (+), *there* iara : iára (+), *they* nakia ∼ nai / naira : hagía (+).

thick tibo- : durú- (−), *thin* uakara / bili- : ibíbię- (?), *think* ikisika : aríta-gua (cp. Lokono *know*), *this* to / li : tóa / léa (+), *thou* bukia ∼ bui : bugúia (+), *three* kabuhin, kabuin, kabun ∼ kabuni- : órua (−), *throw* aboreda : huí- ∼ fuí- / a-có-ra (−), *tie* akora : ágora (+; ágora *bite* and ágora *tie* have as free stems gro and gra respectively), *tongue* -ie : -jéje (+), *tooth* ari : ári (+), *tree* ada : uéue (−), *two* biama, bian ∼ biani- : bíama ∼ bía (+), *up* aiomun : įu (+), *vomit* ereda : euéra / euéreha (?), *walk* aiahadi- / akona : áibuga (?), *wash* asokosa : a-cíba (−), *water* oniabo, uini : dúna (−), *we* uakia ∼ uai, uaiko : uagía (+), *weep* aia- : aiáhua (+), *wet* ioio : dódo- (?), *what* hama : á, ka (?), *when* halika : ída-, áhą- (?), *where* halun : halia (+), *white* harira : harú- (−), *who* hama : ka (−).

wife iraito : úmari (−), *wind* auaduli : garábali (−), *wing* -duna : -aróna (+), *wipe* aburauda : a-rága-ca (−), *with* -oma : -úma (+), *woman* hiaro : hjáru (+), *woods* anaku / konoko : árabu (−), *worm* usehi :

ígei (?), *ye* hukia ~ hui : hugúia (+), *year* uiua : irúmu (−), *yellow* subule- : dumári (−).

2. The foregoing statement is concerned with two problems: (1) the time when the Igneri, or original speakers of what later became the Island-Carib language, split off from the Taino, Lokono, and Goajiro; and (2) the manner in which the Igneri and Taino entered the Antilles. Archeologists have investigated both of these problems and their solutions, if correct, should agree with the ones proposed by Taylor on linguistic grounds. It will be convenient to consider the second problem first.

As the result of stratigraphic excavations in the West Indies and Venezuela, archeologists have been able to establish a chronology of four periods, numbered from I to IV.[12] The first is preceramic and need not concern us here, since the little available information about the historic descendants of the people of this period indicates that they spoke non-Arawakan languages.[13]

Period II is marked by the first appearances of both pottery and agriculture. These clearly entered the Antilles from South America. They seem to have passed from the lower part of the Orinoco Valley onto Trinidad, whence they spread through the Lesser Antilles into Puerto Rico, leaving the rest of the Greater Antilles in its original preceramic status.[14]

The pottery of Period II is strikingly similar wherever it has been found over the vast area extending from Barrancas, just above the Delta of the Orinoco River in Venezuela, through Trinidad and the Lesser Antilles, and into Puerto Rico. Everywhere, this pottery is finely made—more finely than at any later time. Everywhere, too, gracefully outflaring sides and rims with triangular cross sections predominate, and the decoration is characterized by white-on-red painted designs, with the white paint dropping out in some parts of the area during the latter part of the period. The existence of these similarities, extending over such a broad area and always occurring at the time of first appearance of pottery and agriculture, can hardly be explained except as the result of a migration. I have coined the term "white-on-red horizon complex" to refer to the pottery which the migrants seem to have brought with them.[15]

Lovén has identified the pottery of the white-on-red horizon complex as Igneri, and he is followed in this by most archeologists currently working in the Antilles.[16] If it is true, then the Igneri must have entered the Antilles during Period II, and have at that time taken over all of the major islands except Hispaniola, Jamaica, and Cuba.

At the beginning of Period III, pottery related to the white-on-red horizon complex makes its appearance in Hispaniola, Jamaica, and Cuba along with the first evidences of agriculture in those islands.[17] Presumably, this pottery marks the spread of the Igneri through the rest of the Greater Antilles, leaving only isolated pockets of preceramic, non-Arawakan peoples in some of the larger islands.

A little later in the period, the pottery of the Greater Antilles, at least, begins to vary greatly. The previous horizon complex de-

[12] Irving Rouse, The Circum-Caribbean Theory, an Archeological Test, AA 55.188–200 (1953).

[13] Sven Lovén, Origins of the Tainan Culture, West Indies (Göteborg, 1935), pp. 1–6.

[14] Rouse, *op. cit.*, Fig. 2, Profile 1.

[15] Idem, p. 193.

[16] Lovén, *op. cit.*, pp. 271–3; J. A. Cosculluela, Sincronismo de las culturas Indo-Antillanas, Revista de Arqueología y Etnología, II época, 3.27–51 (Habana, 1946), pp. 38–41; J. M. Cruxent, Prehistoric Pottery of Venezuela (MS); etc.

[17] Rouse, *op. cit.*, Fig. 2, footnote 5. The red-painted pottery referred to in this footnote appears to be derived from the ceramics of the latter part of Period II in Puerto Rico. It has now been identified at the beginning of the ceramic sequence in Jamaica and Haiti as well as in the Dominican Republic, to which the footnote refers. See Marian DeWolf, Excavations in Jamaica, American Antiquity 18.230–8 (1953), pp. 237–8.

velops into a series of local styles, differing greatly from island to island.[18] Since there is considerable variation in other aspects of culture as well, it is probable that the inhabitants of the several islands lived in relative isolation. Evidences of large-scale migrations are lacking, although population increases may well have led to expansion of the settlements of individual islands. In Puerto Rico, for example, our site survey has shown that the Indians completed the settlement of the interior of the island at this time.[19]

The local variation and scarcity of evidence of migration have led us to conclude that the Igneri continued to live in both the Lesser and the Greater Antilles through at least the first half of Period III.[20] It does not seem likely that the Taino could have moved into the West Indies at this time.

By Period IV, on the other hand, the Island Carib must have conquered the Lesser Antilles and the Taino made their appearance in the Greater Antilles, for this period extended into historic times, when the Island Carib and the Taino were in possession of those respective areas. As Taylor has noted, the Carib traditions indicate that they moved out from South America and took over the Lesser Antilles from the Igneri. The Taino, on the other hand, had no clear traditions of origin and so we can only infer their movements from archeological and linguistic evidence.

From an archeological standpoint, Period IV was a time of reappearance of similarities in pottery from island to island. These make it possible to distinguish a "modeled-incised horizon complex," extending throughout the Greater Antilles.[21] The roots of this complex go back to Period III, when its designs and techniques of decoration appear gradually in both the Lesser and the Greater Antilles. They are handled differently from island to island, as if they had spread individually and had been adapted separately to each of the local ceramic traditions of Period III. By Period IV in the Greater Antilles, however, they had become standardized and are everywhere found in association with the same shapes, typically bowls with incurving sides.

In view of this gradual development during Period III and of the barrier formed by the Carib occupation of the Lesser Antilles during Period IV, it is difficult to see how the Taino could have migrated from the mainland. Instead, I have elsewhere assumed that the Taino developed from the Igneri in the Greater Antilles.[22] The fact that the deities known as ZEMIS, which were a focal point of Taino culture in the time of Columbus, are first found in the archeological sites of the latter part of Period III and increase in numbers and complexity during Period IV strengthens this interpretation.

Archeology, then, indicates that the Igneri first reached the Lesser and Greater Antilles during Period II and had intensively occupied both regions by Period III. During Period IV they gave way to the Island Carib in the Lesser Antilles and developed into the Taino in the Greater Antilles.

The first part of this reconstruction is consistent with Taylor's linguistic formulation but the second is not. To judge from his linguistic evidence, the Taino language cannot have developed out of the Igneri in the Greater Antilles, as archeologists have assumed. On the other hand, it is difficult to reconcile a migration of the Taino speakers from the mainland with the archeological evidence, as presented above. The two sets of evidence are contradictory, and so we are faced with a dilemma.

Pending the accumulation of further data,

[18] Irving Rouse, Areas and Periods of Culture in the Greater Antilles, SJA 7.248–65 (1951), pp. 256–8.

[19] Irving Rouse, Porto Rican Prehistory, The New York Academy of Sciences, Scientific Survey of Porto Rico and the Virgin Islands 17.305–578 (1952), pp. 566–71.

[20] Idem, Table 1, p. 340.

[21] Rouse, *op. cit.* (footnote 12), p. 193.

[22] Rouse, *op. cit.* (footnote 19), pp. 361–2.

I do not know of any satisfactory way to resolve this dilemma, but three possible solutions occur to me:

(1) The most obvious possibility is that the linguists and the archeologists are not talking about the same thing when they speak of Igneri and Taino. Linguists apply these terms to languages but archeologists must of necessity define them in cultural terms, with emphasis on pottery. It has long been customary to assume that language and culture are directly correlated in the Caribbean area but, as I have noted elsewhere,[23] this is not necessarily true. The fact that the Island Carib spoke an Arawakan language but had a typically Carib culture is a case in point. Can it be that the languages and cultures of the West Indies had entirely separate histories and that, therefore, one should not attempt to correlate them, as is being done here?

To answer this question in the affirmative would, in my opinion, be unwarranted. Certainly, the appearance of agriculture and of white-on-red pottery during Period II must have been accompanied by the introduction of a new language, presumably Igneri. On the other hand, one might assume that the Taino language was subsequently introduced into the Antilles without such a marked change in the culture that it would show up in the archeological record. This would provide one possible solution to the dilemma.

(2) As a second, related possibility, it can be suggested that we archeologists may have been too cautious in inferring migrations from our remains. If it is true that changes in language do not always correspond exactly to changes in culture, then perhaps we should look for less pronounced innovations in culture with which to correlate the migration of the Taino into the Antilles. Two such innovations are perhaps worth mentioning:

(a) Most of the white-on-red pottery of Period II lacks any great amount of incision and modeling, but there is a sub-complex which has these characteristics—particularly finely incised crosshatched designs.[24] The exact relationship of the sub-complex to the more typical white-on-red pottery is still uncertain, but there is reason to believe that it may have originated along the north coast of Venezuela, whereas the more typical pottery appears first—so far as we now know—in the lower part of the Orinoco Valley. Could these two forms of the white-on-red horizon complex have been brought into the Antilles by separate migrations, one of the Igneri and the other of the Taino?

(b) If, instead, only the Igneri pottery reached the West Indies during Period II, along with the white-on-red pottery, then the Taino would have had to follow during Period III, despite the apparent local diversity at that time. In such an event, the first appearance of elements of the subsequent modeled-incised horizon complex and of the worship of ZEMIS might be considered a marker of the Taino migration.

(3) Taylor has suggested as an alternative hypothesis to (1) and (2) that the Taino may have entered the Antilles before the Igneri, in which case it would have been the Taino who introduced agriculture and white-on-red pottery during Period II. The Igneri would then have come in later, either during the latter part of Period II, as in (2, a) above, or during Period III, as in (2, b). It is even possible—although not so likely—that the Island Carib acquired the Igneri language on the mainland and themselves introduced it into the Antilles during Period IV.

[23] Irving Rouse, Guianas, Programa de Historia de América I, 7 (Mexico City, 1953), pp. 24–6.

[24] This sub-complex has never been defined as such, but the Cedros style in Trinidad may be cited as an example. Irving Rouse, Prehistory of Trinidad in Relation to Adjacent Areas, Man 47.93–8 (1947), pp. 93–4. For its occurrence in Venezuela, see Cruxent, *op. cit.* It is also common in the Lesser Antilles and has recently been isolated by Ricardo E. Alegria (personal communication) in Puerto Rico, where its relation to my Cuevas style is uncertain.

Taylor has expressed a preference for the second of these three alternatives. If I had to make a choice, however, I would prefer (3). I cannot see the Taino having by-passed the Lesser Antilles to impose their language and (presumably) their culture on the Greater Antilles when, according to the archeology, both regions were just as intensively occupied by the earlier, supposedly Igneri people. The Igneri—if they were the first inhabitants—must have been in enormously greater numbers on the larger islands of the Greater Antilles than on the tiny ones of the Lesser Antilles and so should have been much easier for the Taino to assimilate in the Lesser Antilles; yet, hypothesis (2) assumes that they did not do so.

A second reason for my preferring hypothesis (3) is that it better fits the culture of the Taino, as known ethnologically. These Indians are reported to have been peaceful—their weapons, for example, were "exceedingly under-developed,"[25]—and therefore it is difficult for me to envisage them in the role of conquerors, as required by hypothesis (2). Hypothesis (3), on the other hand, assumes that the Igneri instead of the Taino were warlike, and this better fits what we know of the Igneri who survived the Carib migration on Trinidad—they did have good weapons, for one thing.[25]

Taylor, to be sure, argues that the Taino would not have had to conquer the Igneri but could have settled peacefully among them, as the few Negro escapees subsequently did among the Island Carib. This, however, seems to me a false analogy, since the Negroes and the Carib were united by a common bond of hatred against the Europeans, which the Taino and Igneri could not have shared. Moreover, the Negroes were acting as individuals, whereas the Taino presumably came into the Antilles as a well organized social and political group. For such a group, conquest should have been a necessity, unless the Taino were able to win over the Igneri by superior prestige of some sort, which does not seem likely during Period III, when both groups were apparently on the same level of cultural development.

Finally, I prefer hypothesis (3) because it can be more easily reconciled with the archeological record. It accounts for the settlement of the Greater Antilles by means of only a single migration, that of the Taino during Period II, and so does not do violence to the evidence for lack of migration in that region during Periods III and IV. This hypothesis has the advantage of restricting the second, Igneri migration to the Lesser Antilles, where the archeology is as yet so poorly known that it cannot be said for sure whether one or two migrations took place.

There remains the problem of the date at which the Igneri split off from the Taino, Lokono, and Goajiro. Our uncertainty about hypotheses (1), (2), and (3) above should not affect the handling of this problem, since the split presumably took place on the mainland, before either the Igneri or the Taino entered the Antilles. The time of Period II, therefore, may be used as the point of comparison with Taylor's lexicostatistical date, no matter whether it was the Igneri or the Taino who reached the Antilles at that time.

In connection with excavations in Puerto Rico, I have obtained a series of dates by determining the amount of time required for the deposition of the refuse attributable to the various prehistoric periods.[26] I first calculated the yearly rate of deposition of refuse in the historic sites and then divided this into the average depth of refuse in the sites of each prehistoric period to determine the duration of each. Adding the latter figures together, I obtained a date of *ca.* 930 A.D. for the introduction of the white-on-red pottery at the beginning of Period II. This is a crude method of calculation, if only because it rests on the unlikely assumption that the rate of deposition was the same

[25] Lovén, *op. cit.*, pp. 440–1.

[26] Rouse, *op. cit.* (footnote 19), pp. 564–6.

during Periods II, III, and IV as in historic time, but it was the best that could be done under the circumstances.

Elsewhere,[27] I have commented that subsequent work is likely to lengthen this date rather than to shorten it. I am unable, however, to agree with Willey[28] when he states that the date is "impossibly late," nor can I accept the estimate of Father Pinchon[29] that the Igneri reached Martinique in the Lesser Antilles "au début de l'ère chrétienne," since neither author gives a basis for his opinion. Pending the presentation of evidence in favor of an earlier date, I do not think that my Puerto Rican calculation needs to be pushed back much before 600 A.D.

Archeology, then, gives an approximate date of 600 to 930 A.D. for the presumed arrival of the Igneri or Taino in Puerto Rico. This is to be compared with Taylor's date of *ca.* 1 to 300 A.D. for the differentiation of the Igneri and Taino languages on the mainland. Taking the latest value for the latter date and the earliest value for the former, we obtain a difference of 300 years between the two. This does not seem to be an unreasonable length of time to allow for the separation of the Igneri and Taino languages on the mainland and the spread of one of them through the Lesser Antilles into Puerto Rico.

A further check on the validity of the linguistic date will soon be provided by the method of radiocarbon analysis. J. M. Cruxent, of the Museo de Ciencias Naturales in Caracas, and I have obtained a number of samples of charcoal from deposits with white-on-red pottery at the Saladero site on the lower Orinoco, which are now undergoing analysis at the Geochronometric Laboratory in Yale University.[30] They should provide a more reliable archeological date and one which is more comparable to the linguistic date, since they come from the mainland where the separation of Igneri and Taino language is supposed to have taken place. Definitive conclusions cannot be drawn until the forthcoming radiocarbon dates are obtained, but meanwhile such archeological evidence as we have tends to support Taylor's linguistic date.

[27] Rouse, *op. cit.* (footnote 23), pp. 17–19.

[28] Gordon R. Willey, Review of "Porto Rican Prehistory," AA 56.138–41 (1954).

[29] Père Robert Pinchon, Introduction a l'archéologie martiniquaise, JSAP 41.305–52 (1952), p. 347.

[30] Cruxent, *op. cit.*

A NOTE ON GLOTTOCHRONOLOGY

GORDON H. FAIRBANKS
CORNELL UNIVERSITY

0. Introduction
1. Problem
2. Method
3. Data
4. Conclusions

0. There has been considerable interest in the application of statistical techniques to the study of language vocabulary, variously called glottochronology and lexico-statistical dating. It was first established on the basis of counts from 13 languages that 'on the average about 81% of the basic-root-morphemes of a language will survive as cognates after 1000 years.'[1] This implies that two related languages, 1000 years after they have split off from the parent language, will have 66% of cognates from the basic morpheme list, that is, each will retain 81% of the basic list and the overlapping of cognates will be 81% of 81%, or 66%. It is this principle that is used by Swadesh in dating Salish relationships,[2] although, since he does not use the same list that was used to establish the retention rate of 81%, he can establish relative relationships but not an absolute chronology. Swadesh also uses this principle, along with other data, to confirm the existence of a genetic unity for Salish, Chimakuan and Wakashan.[3] Taylor uses this principle to establish the time of split of two Carib dialects.[4] This use of glottochronology implies the determination of cognates, which in turn implies the complete, or almost complete, determination of phonological correspondences between the languages involved. In the case of the Salishan languages there is as great a time depth, 11% of common cognates in Bella Coola and Kalispel, as in the Indo-European languages. Although Swadesh' article reads as if the phonological correspondences in Salishan were well established, it might not be unfair to assume a greater degree of indeterminacy in Salishan than in Indo-European. Swadesh estimates a time depth of 55 centuries for Salishan. For Mosan[5] Swadesh estimates 90 centuries, and here we can certainly assume a much greater degree of indeterminacy in the establishment of phonological correspondences. This leads us to wonder whether some of the morphemes assumed by Swadesh to have a common origin in Mosan may not be morphemes that show apparent agreement by chance. Swadesh is himself concerned about this possibility and makes a rough calculation of the probability of chance correspondences to be expected between Chehalis and Kwakiutl. Assuming there are in the first language 28 phonemes that may occur in initial position and that in second position 29 may occur, he calculates the chance factor as $\frac{1}{28} \times \frac{1}{29} = \frac{1}{812}$. He then increases this factor to $\frac{1}{200}$ to allow for 'unequal frequencies and for phonological flexibility (e.g. comparing glottalized with non-glottalized consonants).'[6] This allowance may not be sufficient. If we are comparing morphemes in two languages and are not sure of the phonological correspondences, then we are not sure whether we are counting cognates or just similar items. If we are counting similar items that may be cognate, then we would accept them as similar not

[1] M. Swadesh, Lexico-statistic dating of prehistoric ethnic contacts, APS-P 96.452-63 (1952) and Robert B. Lees, The basis of Glottchronology, Lg. 29.113-27 (1953).

[2] M. Swadesh, Salish internal relationships, IJAL 16.157-67 (1950).

[3] M. Swadesh, Mosan I: A problem of remote common origin, IJAL 19.26-44 (1953).

[4] Douglas Taylor, Sameness and difference in two Island Carib dialects, IJAL 18.223-30 (1952).

[5] M. Swadesh, Mosan I, IJAL 19.26 (1953).

[6] Ibid. 43-4.

only if they show identity of consonants, i.e. t − t, but also if they differ in a reasonable manner, e.g. t − d, t − θ, t − s, t − c, etc. Assuming 28 possible initial consonants, the probability that a second language will have a morpheme beginning with a t corresponding to a morpheme in the first language that begins with a t will be $\frac{1}{28}$. The probability that the morpheme in the second language begins with t, d, θ, s or c, not to mention other possibilities, is greatly increased, maybe to as much as $\frac{1}{5}$. If this is a more reasonable rough calculation than that of Swadesh, then we must revise his figure from $\frac{1}{200}$ to, say, $\frac{1}{5} \times \frac{1}{5} = \frac{1}{25}$. In Chehalis and Kwakiutl using the chance factor of Swadesh we would expect 5 out of 1000 items to appear similar by chance, but using our revised figure we would expect 40 similar chance items. Swadesh says that Boas found 48 apparent cognates out of 1000 items, which is not very much more than would be expected by chance, but Swadesh has found 40 more, which is twice the number expected by chance, certainly a reasonable margin.

1. The doubt raised in the previous section as to whether Swadesh might have been counting not cognates but similar forms in some of his Mosan work introduces a new problem. Is it valid to count similar morphemes in two languages, leaving aside all consideration of cognates, and use this as a method of determining genetic relationship? Swadesh[7] has done this to some extent in an unpublished paper. William Bright[8] has applied this method to the Hokan languages with interesting results. The idea of such a comparison is a shock to a comparativist. We are apt to think at first of the individual example. The form bad in English and Persian is a well-known example of a form that is similar in the two languages but not cognate. From this we might draw the conclusion that forms that are very similar in two languages as distantly related as are English and Persian are certain not to be cognate. This is supported by such dissimilar cognates as Eng. comb /kowm/ and Russ. zub, or Eng. two and Armenian yerk'u. There is also evidence to the contrary: Eng. water and Hittite watar, which are both cognate and similar. We have no data, to my knowledge, on the various problems raised by this discussion: How many cognates and how many similar items do two languages show? How many items are cognate, but not similar? How many are similar but not cognate? Is there any correlation between the number cognate and the number similar? The rest of this paper will present some data on these problems which, it is hoped, will be of interest.

2. The basic morpheme list of 215 items compiled by Swadesh[9] was used. The languages used were English, French, German, Czech, Russian and Armenian (Modern East Armenian). Two counts were taken of various pairs of these languages. The first count was of the cognate root morphemes. This count will be called the cognate count, and the method the cognate method. The second count was of similar forms, whether cognate or not. It is apt to be difficult to determine similarity without any special criteria for similarity particularly if you already know whether items are cognate. Also, if specific criteria are used, it is possible for the investigation to be duplicated by others. For these reasons I have set up certain criteria to determine similarity, criteria which are similar to those used by Bright in his investigation of the Hokan languages. This count is called the inspection count, and the method the inspection method. Where possible I select two consonants from

[7] M. Swadesh in a paper presented at the Linguistic Institute held in Chicago in the summer of 1954. The paper was much greater in scope than might be suggested by my remarks. What is pertinent here is that in the paper Swadesh compared similar forms, albeit from a large number of languages.

[8] William Bright presented this paper in a seminar on Middle American Languages directed by Norman McQuown at the Linguistic Institute held in Chicago in the summer of 1954.

[9] M. Swadesh, APS-P 96.456-7 (1952).

what might be considered, on the basis of inspection, to be the root morpheme. Vowels I have generally neglected except for an occasional instance where I was in doubt. I consider two consonants to form a similarity correspondence (contrasted with the usual phonological correspondence) if they differ by one feature or component, thus Eng. do and Ger. tun agree with respect to the first consonant since Eng. d and Ger. t differ in the feature of voicing. The aspiration of Ger. t is not considered because in neither language is this a significant feature. Likewise Eng. this, Ger. dies agree since Eng. đ and Ger. d differ by the feature of spirantization. I admit as a similarity correspondence two consonants that differ by two features if the other consonants under consideration do not differ at all and if there is no great excess of unexplained material in either item, thus Eng. thick, Ger. dick, Eng. thin, Ger. dünn, although the first consonants differ both in voicing and spirantization. I have not admitted Eng. brother, Arm. yexp'ayr, where we could set up the correspondence, Eng. b − r, Arm. p' − r, because there is too much unexplained material, yex in Armenian and đ and an extra r in English. I have not under any conditions allowed p − k, which might be said to differ by two degrees of position, although this restriction may be a prejudice based on some knowledge of Indo-European.

The choice of languages is partly dictated by those immediately available to me,[10] but in part the choice was meant to include languages that would show varying percentages of cognates. English and German, Czech and Russian, were chosen as languages with a high percentage of cognates. Normally there would be no point in using the inspection method on languages with such a high percentage of cognates since it would be immediately obvious that they were related languages and the phonological correspondences could be set up with a high degree of accuracy. These languages should be of interest, however, for some of the other questions posed in section **1.** The other languages were chosen partly as examples of varying percentages of cognates, although the percentage differences were less than might have been expected, and partly to get one language from different branches of Indo-European, Germanic, Italic, Slavic and Armenian.

[10] These lists were compiled by me in connection with an entirely different glottochronologic study on which I am collaborating with Eric Hamp and Herbert Paper.

3. The results of the two counts were as follows:

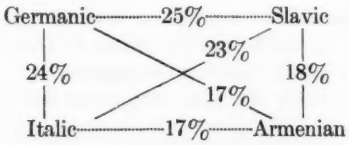
	Cognate Count		Inspection Count	
	No. of cognates	% of cognates	No. of similar items	% of similar items
Czech-Russian	160	74%	162	75%
English-German[11]	128	60%	129	60%
English-Czech[12]	54	25%	39	18%
English-French[12]	51	24%	53	25%
French-Russian	49	23%	38	18%
Russian-Armenian	38	18%	36	17%
English-Armenian	36	17%	32	15%
French-Armenian	36	17%	36	17%

[11] My figures for cognates may be compared with those of R. B. Lees, Lg. 29.122 (1953), where he finds 124 cognates, 58.5%, in a total of 212 items. I have used 215 items. I have considered Eng. cook, Ger. kochen as cognate in spite of the fact that this is borrowed from Italic because it must have been borrowed before the High German consonant shift and is stated by Götze to have been borrowed at the latest in the fourth century, cf. koch in Götze, Etymologisches Wörterbuch der deutschen Sprache, Berlin (1934).

[12] It may be noted here that I compared Czech with English, but Russian with French and Armenian. This was not originally intended, but should not be significant since either Czech or Russian can serve as an example of a Slavic language and should not affect the data.

[13] Again my figures for cognates can be compared with those of R. B. Lees, Lg. 29.123 (1953), where he quotes 56 cognates, 27.7%, out of 202 items. This difference is a little more significant than our difference in the Eng-Ger list, but I cannot account for it.

The mutual relationships of these branches can probably be better illustrated by the following chart:



Although this relationship is not particularly pertinent to the present paper, it may be noted that Germanic, Italic and Slavic are shown to be related at approximately the same time depth, whereas Armenian is more distantly related to the other three and the relationship is consistent.

A count was also taken of those items which were similar but not cognate and of those that were cognate but not similar.

	Cognate but not similar	Similar but not cognate
Cz-Rus	1	3
Eng-Ger	4	5
Eng-Cz	24	9
Eng-Fr	17	19
Fr-Rus	22	10
Rus-Arm	14	13
Eng-Arm	17	13
Fr-Arm	13	13

4. There is enough consistency in the data presented above so that it is worthwhile to draw a few conclusions from them. It should, of course, be recognized that the generalizations are based on a small amount of data and therefore need further confirmation or disproof.

The language pairs that have a high percentage of cognates, Eng-Ger 60%, and Cz-Rus 74%, show a very high degree of correlation between the forms that are cognate and those that are similar. It is only rarely that a cognate item is not also similar. In Cz-Rus there is only one form, Cz protože, Rus. potomučto because, and this might be considered similar. In Eng-Ger there are four, Eng. how, seed, tooth, who, Ger. wie, Same, Zahn, wer, and tooth might be considered similar to Zahn if we relaxed the criteria only very slightly.

The Eng-Cz count with 25% cognates and the Fr-Rus count with 23% cognates agree in both showing a noticeably smaller percentage of similar forms, both 18%. At first glance the Eng-Fr count seems to disagree with this, but of the 53 similar items 11 are borrowings either from French into English (Eng. animal, flower, lake, mountain person, push, turn, vomit, dance) or from Germanic into French (Fr. flotter, gratter, Eng. float, scratch).[14] This would reduce the number of similar items to 42, or 20%, which would then agree with the other two counts. The languages in this 25% cognate range all show a much greater number of items that are cognate but not similar than those in the high range. Very nearly half of the cognate items are not similar, although the Eng-Fr counts deviate somewhat in having only one-third of the cognate items dissimilar. The counts in this range also agree in having about the same number of similar items that are not cognate, Eng-Cz 9, Fr-Rus 10, and Eng-Fr 8, if we discount the 11 borrowed forms. It would seem that languages in this range will have a significantly smaller percentage of similar forms than cognate forms, that a fairly high percentage of the cognates, 33%–45%, will be dissimilar, which will be partially but not completely offset by approximately 25% of similar items not being cognate.

The last three counts with 17% or 18% of cognates agree in having just about the same percentage of similar forms. These counts agree with those in the 25% range in having from 30% to almost 50% of the cognates dissimilar. They differ from the preceding group in having the same proportion of similar non-cognate items as cognate non-similar items.

[14] Apart from the English the item animal would have to be eliminated since it is a learned borrowing in French also. The item dance has been borrowed back and forth across the Rhine and should be eliminated on several scores.

I have no data on languages that have over 25% and less than 60% of cognates, but it would be surprising if such data would show significantly different results. I also have no data on languages that show a smaller percentage of cognates than 17%. This is more serious. Before we can make any sort of statement about the possibility of determining language relationship on the basis of a count of similar items when the percentage of similar forms is less than 15% we need data of the type presented here for languages with varying lower percentages.

The data presented here also confirm the suggestion made above that the 1 in 200 chance similar items assumed by Swadesh is much too low. I suggested that 1 in 25 might be more reasonable, and the number of forms that are similar but not cognate, that is, forms that are similar by chance, is even higher than that estimate.

Probably the most interesting fact to be gleaned from these data is that there is a consistent relationship between the cognate count and the similar count. The data is insufficient to draw any far-reaching conclusions, but suggests that further investigation would be justified.

TOWARDS GREATER ACCURACY IN LEXICOSTATISTIC DATING[1]

MORRIS SWADESH
DENVER, COLORADO

1. Introduction
2. The method
3. Accuracy
4. The test vocabulary
4.1. Universality
4.2. Cultural implications
4.3. Simplicity
4.4. Interlingual ambiguity and semantic shading
4.5. Potential duplication
4.6. Identical roots
4 7. Sound imitation
4.8. Form words
5. Item persistence
6. Control procedure
6.1. Overlapping histories
6.2. Hidden divergence
6.3. Time depth weighting
6.4. Dating of samples
6.5. Scoring
6.6. Modifying factors
7. Provisional evaluation

1. The last several years have seen a number of relatively successful experiments by various scholars in estimating time depth by counting cognates in test vocabularies. This development is of great interest for the reconstruction of prehistory because it offers a time dimension for ethnic contacts shown by linguistic relationships. In the light of historically known locations of peoples, it gives evidence regarding their prehistoric location and movements at given periods of the past. This data can be correlated with the indications of archeology, comparative ethnography, geography of loan-words, geographic distribution of phonetic and morphologic features, and many other kinds of evidence. It frequently

[1] Articles bearing on lexicostatistic theory and method:

WALTER W. ARNDT, Germanic Dialect Evolution in Lexico-Statistic Time Perspective, University of North Carolina Doctoral Dissertation, 1955.

WILLIAM W. ELMENDORF, Word Taboo and Lexical Change in Coast Salish, IJAL 17.205–8 (1951).

R. FOX, W. SIBLEY AND F. EGGAN, A Preliminary Glottochronology for Northern Luzon, Proceedings of the Eighth Pacific Science Congress, Manila.

C. F. HOCKETT, Linguistic Time Perspective and its Anthropological Uses, IJAL 19.146–52 (1953).

DAVID I. HIRSCH, Glottochronology and Eskimo and Eskimo-Aleut Prehistory, AA 56.825–38 (1954).

G. EVELYN HUTCHINSON, The Half Life of a Word (Marginalia), American Scientist 41.633–34 (1953).

ROBERT B. LEES, The Basis of Glottochronology, Language 29.113–27 (1953).

——— Cognate Word Counts, unpublished. (Scored test lists used in the preceding.)

MORRIS SWADESH, Salish Internal Relationships, IJAL 16.157–67 (1950).

——— Diffusional Cumulation and Archaic Residue as Historical Explanations, Southwestern Journal of Anthropology 7.1–21 (1951).

——— Mosan I: A Problem in Remote Common Origin, IJAL 19.26–44 (1953).

——— Lexico-Statistic Dating of Prehistoric Ethnic Contacts, Proceedings of the American Philosophical Society 96.452–63 (1952).

——— Archeological and Linguistic Chronology of Indo-European, AA 55.349–52 (1953).

——— The Language of the Archeologic Huastecs, Notes on Middle American Archaeology and Ethnology 4.223–27 (1953).

——— Perspectives and Problems of Amerindian Comparative Linguistics, Word 10.306–32 (1954).

——— Fechas Glotocronológicas importantes para la Prehistoria Nahua, Revista Mexicana de Estudios Antropológicos, in press.

M. SWADESH, GEORGE I. QUIMBY, HENRY B. COLLINS, EMIL W. HAURY, GORDON F. EKHOLM AND FRED EGGAN, Symposium, Time Depths of American Linguistic Groupings and Comments, 56.361–77 (1954).

gives important clues for connecting prehistoric archeologic sites with historically known peoples.

Within the field of linguistics itself, lexicostatistics brings to light a previously undiscovered reality—the tendency of vocabulary to change at a uniform rate. Since lexicostatistic work thus far has been keyed to the estimation of time-depths and has purposely sought to use those lexical elements which are least subject to cultural influences, it follows that only a fraction of the total phenomena have been studied. Eventually it will be desirable to study rates of change in the various types and levels of less stable vocabulary and to develop a complete theory of the factors affecting the rate of vocabulary change.

Despite the broad promise of an added instrument of prehistoric research and a further key to the understanding of language as a social phenomenon, lexicostatistics has thus far operated, as it were, on a shoe-string of basic research. What has been called the 'constant', but might be better called the 'index', of lexical retention is based on a partial study of only 13 control languages, a most valuable initial contribution made by R. B. Lees, working in communication with the present author and incorporating data supplied by several other scholars. The various experiments in the application of lexicostatistics to prehistory and certain studies attempting to verify the validity of the index and to check special factors, all have contributed to a better grasp both of the methodology and the theory upon which it rests. Nevertheless, the need remains for a large-scale control study to include perhaps a hundred historically known cases.

If there were a guarantee that the necessary pilot study would be undertaken and carried to fruition within a few years, it might be well to avoid piecemeal research and even to suspend further time-depth estimates temporarily. However, since there is at present no such prospect, it is well to continue experimentation in the hope that gradually accumulating knowledge will eventually perfect the practice and theory of lexicostatistics. The present contribution, along these lines, is a very modest one. It presents no new control languages and even drops, for methodological reasons, some of those listed by Lees. It attempts a further analysis of the old word lists, and to utilize the many criticisms and suggestions coming from colleagues in publications, discussion and correspondence. Its chief value is probably that of exploring the means available for a more thorough-going improvement.

2. In the methodology of lexicostatistic dating, a set list of test items is used. They are expressed in some convenient translation language, such as English or Spanish or any other, but it is understood that a certain meaning is intended for each item independent of any semantic peculiarities in the translation language. For instance, *tongue* refers to the body-part, not to a language and not to a part of a wagon. For each test item the common, everyday equivalent is listed for each language. For example, if we were comparing French and Spanish, we would set down *langue* for the first and *lengua* for the second as the equivalents of *tongue*; for *head*, it would be *tête* and *cabeza* respectively. Ordinarily only one equivalent is taken for each test item, more than one only when there is no way of choosing among the common equivalents. The lists for the two languages are then set parallel and scored for cognates and non-cognates; French-Spanish *langue-lengua* are scored plus, *tête-cabeza* are minus. The cognates are counted and the percentage of cognates is determined with respect to the total of cognate and non-cognate pairs. The total is of course the number of items in the test list, unless some of them cannot be supplied or cannot be scored as to cognacy, say because of unresolved problems of sound correspondence. The most general

equation for converting the cognate percentage into an estimate of units of time is:

$$t_1 + t_2 = \log C \div \log r$$

In this, t_1 and t_2 relate to two separate dialects or languages developed out of what was originally one language; t_1 is the time, in millenia (or other convenient time unit), that separates one of the languages from the common period, and t_2 is the time of the second. C is the percent of cognate elements found on comparing the two test lists. And r is the retention index, that is the expected percentage of elements retained in languages after one unit of time, as established by control studies.

In unilinear problems, comparing two stages of the same language, one of the language forms is contemporary with the common period, so that t_1 is zero, and $0 + t_2 = \log C \div \log r$, or $t_2 = \log C \div \log r$. If the two language forms are both later stages, we are dealing with a problem of divergence. If they are contemporaneous, then $t_1 = t^2$; hence:

$$2t = \log C \div \log r$$

$$\text{or} \quad t = \log C \div 2 \log r$$

$$\text{or} \quad t = \log C \div \log r^2$$

If t_1 and t_2 are unequal, it is possible to calculate one of them if we know the other ($t_1 + m = \log C \div \log r$, with m being a known length of time) or if we know the difference between the two ($t_1 + t_1 + d = \log C \div \log r$, or $2t_1 + d = \log C \div \log r$).

The simple relation between cognate percent and time applies under conditions of full or undiminished divergence, that is, when there are no contact influences between the two diverging dialects to keep their lexicons from going their separate ways. To the extent that there is contact, the rate of divergence will be slowed. To allow for this, we may define divergence as $\bar{s}t$, that is average degree of separation multiplied by the time, and then revise the equation to read:

$$t = \log C \div \bar{s} \log r^2$$

$$\text{or} \quad \min t = \log C \div \log r^2$$

In the first modification, $\bar{s}$ is a quantity never larger than 1, being equal to 1 under circumstances of undiminished divergence and to a smaller or greater fraction in inverse proportion to the degree in which contact influence interferes with divergence. The equation with $\bar{s}$ is valuable mostly to help conceptualize the effect, but there is no way of calculating $\bar{s}$ until one knows t. One therefore must begin by calculating the time periods among all possible pairs of languages in a genetic grouping, while remembering that the actual time may be greater than that indicated; this is the meaning of the second modification of the equation, using min t. The procedure is thereafter reconstructive, using various guiding rules, of which these two may be the most important: (1) Where there is evidence of complete geographic separation between two languages, it follows that the divergence was unhindered by contact influences. (2) In a set of indicated time-depths among the languages of a genetic grouping, the greatest is most likely to correspond to the actual time since the common period.

The index of retention, as calculated by Lees on the basis of 13 control languages over time-spans of 1000 to 2200 years, is 80.48% ± 1.76% for 1000 years, the limits of error being calculated as "the $9/_{10}$-error in the mean, using small-sample methods." The retention rate used in most studies is taken as 80.5 or 81 pct, and r^2 (unit index of cognate retention) is accordingly 65% or 66% per 1000 years.

3. There is already sufficient evidence that lexicostatistically estimated time depths are to some degree approximately related to actual time. This is shown by a number of correlations, including several historical dates and a few archeological ones, including some obtained by the carbon method. And there is further evidence of general cor-

rectness in the inner consistency of lexicostatistic dates found among the languages of a family or stock. Without claiming for lexicostatistics the accuracy of a precision instrument, there can nevertheless be no doubt that it has considerable approximate validity and is already capable of serving as a helpful tool for the reconstruction of prehistory.

At the same time, there is every reason to believe that lexicostatistics has not yet reached or even approached its maximum potentialities. This follows from the limited basic research underlying both the index of retention (previously called the 'constant') and the diagnostic list. And various scholars who have studied or used the method have pointed out specific weaknesses most of which appear capable of correction. Although a final revision of the method is not possible now, we can correct many of the minor deficiencies and explore techniques for still further improvement. The main effort in the present paper is with respect to the diagnostic list. Certain changes are attempted and their effect on accuracy quantitatively evaluated. As will be seen, the result is markedly favorable.

4. "Suitable items for a test list must be universal and non-cultural . . . Moreover, they must be easily identifiable broad concepts, which can be matched with simple terms in most languages." (Lexico-Statistic Dating 457.) These criteria, recognized from the first in lexicostatistic research, were not consistently followed in setting up the original test list. In fact, fifteen of 215 items used by Lees in his pilot study, were already recommended for discard in Lexicostatistic Dating (published before though written after Lees' research). In addition to these criteria, some others came to light as the result of experience, including as negative features to be avoided: potential duplication (noted by Douglas Taylor), identical roots, sound imitation, semantic shading.

It was early recognized that: "It is doubtless possible to devise a better test list than the present one, but from the author's own experiments along these lines it is not too easy." (Lexico-Statistic Dating 457). The first list, in its earlier and modified form, contained about 200 items, and the author hoped at one time to enlarge it in order to gain increased statistical accuracy, however only a handful of really sound new items were found, while on the other hand defects in the old list were repeatedly made evident. The only solution appears to be a drastic weeding out of the list, in the realization that quality is at least as important as quantity. The new list of 100 items includes 92 from the old list, those starred in Table 2, plus eight new ones: say, moon, round, full, knee, claw, horn, breast. Even the new list has defects, but they are relatively mild and few in number.

4.1. If lexicostatistics is to be widely applied, it is convenient to have test items which can be supplied everywhere or nearly so. To carry this criterion out more fully, we now drop the words ice, snow, freeze, snake, sea.

A contrary suggestion has been to develop variants of the test list for different large areas of the world. This would require great care both in the selection of the local items and in determining variations in the retention rates connected with each list.

4.2. The reason culture terms have to be avoided is that their retention or loss is too closely correlated with fluctuations in the cultural situation to serve as an index of the passage of time. It was at first thought that some culture concepts, artifacts and activities were so ancient and so widespread as to exempt them from the general defect of culture terms, and the following were therefore included in the old list (first revision): father, mother, husband, wife, spear, rope, stab (pierce), sew, hunt, play, fight. How-

ever, practically all of these items are unsatisfactory for other reasons, such as non-universality, duplication, ambiguity, sound imitation (or something related to it).

Numerals have to be regarded as cultural. In Mexico we find Indian languages which have taken over the Spanish numerals and the same thing has happened in Japanese with respect to Chinese. This shows the tie-up between numbers and the patterns of commerce and technology. The high persistence of numerals among the control languages—various European languages, Egyptian and Chinese—shows a heightened tenacity probably connected with the place of counting in these cultures. On the other hand there are languages said to lack numbers above *two* or which use compounds for *three* and over. The new word list retains *one* and *two* but drops the others.

There are some natural objects whose cultural role is so strong that they take on the characteristics of culture terms. Examples from the old list are: *salt*, a trade article; *flower* and *fruit*, since cultured varieties present something culturally new; *ice*, strictly a manufactured article in some areas. One of the characteristics of culture terms is the readiness with which they may be borrowed along with the objects or along with new types of objects already known. And this feature helps identify this kind of secondary culture term. Of course, any word whatsoever may take on cultural implications. We can only avoid those which are potentially culture terms in a relatively high degree.

Names of animal species for the most part cannot be used in a test vocabulary both because of non-universality and because they pass with ease from language to language, particularly when a people migrate into an area of new species or of new varieties of the old. The only animal names retained in the revised test list are: *bird fish, dog, louse.*

4.3. All specialized concepts have to be avoided, since they are too frequently expressed by phrases rather than simple words.

4.4. Certain activities like *cut, pull, dig, squeeze* have two characteristics which make them unsatisfactory as test items. One is that it is frequently hard to match them in given languages, and the other is that they belong to a distinctly lower level of stability than the bulk of our test items. The two qualities are probably interrelated. Contrast the stability of *drink, eat, sleep* (cp. German *trinken, essen, schlafen*) with the instability of *cut, pull, dig* (cp. *schneiden, ziehen, graben*).

There are also interlingually ambiguous and unstable object words. Even body parts include such terms as *leg, back, guts*, whose reference is fairly inclusive and shades with that of other words. The leg includes the foot, ankle, calf, shin, knee, thigh. Many languages have a single word for foot and leg. Others make a distinction between the lower and upper leg. Little wonder therefore that *leg* turns out to be one of the least stable items in the list.

4.5. Just as it would be incorrect to have the same word twice or more times in a test list, so is it also desirable to avoid test items which tend to be synonymous. In comparing two languages both of which have *wife* identical with *woman*, one would find that either both items or neither are cognates. To avoid the overweighting which results, the simplest method is to eliminate *wife* from the test. In the same way, we drop *river, lake* and *sea* because of *water*; *far* because of *long*; *short, thin* and *near* because of *small*; *dust* because of *earth*; *fog* because of *cloud*; *leg* because of *foot*; *they* and *he* because of *that*. It would be difficult to remove every last potential duplication, and a few cases of possible duplication still remain in the revised list: *water-rain, skin-bark, big-long, how-what, who-what,*

this-that. It is hoped that the number of languages having any duplications at all will be greatly reduced and that very few indeed will have more than one.

Another approach to the problem of potential duplication would be to leave the cases in but to mark them specifically, so that they may be scored only when the compared languages do not involve duplication. This would enable us to add some otherwise excellent test items—including *lip* (tending somewhat to duplicate *mouth* or to be expressed as a compound *mouth-skin*) and *arm*, in addition to some of those already mentioned—but would complicate the scoring by virtue of the need to check each comparison for synonyms.

4.6. Very similar to the phenomenon just discussed is the use of identical roots in sets of words like *this-that*, *who-what-when-where-how*, *I-we*, *thou-ye*, *die-kill* (the latter a causative formation). The revised list therefore drops *when*, *where*, *how*, *ye* but it keeps the rest of the words as minor risks. When a language has two equivalents for *we*, inclusive and exclusive, the former is to be used because it is less likely to repeat the root of the first person singular.

4.7. Words for *blow* tend to have labial or sibilant consonants or both, along with rounded vowels. A good illustration is Spanish soplar, which has all the imitative sounds, assembled in the process of its development from Latin sub-flare. Even though the vowel and the sibilant come from an affix (sub- *under*) which originally had nothing to do with blowing, it seems very likely that they became associated with the old root (fl-) largely because of their symbolism. Further potential imitatives eliminated from the test list are: *breathe*, *laugh*, *puke*, *scratch*, *laugh*, *cry*. The objection to sound imitative tendencies in the test items is twofold, that their stability is somewhat unpredictable and that in scoring it is often difficult to decide whether one is dealing with cognates or accidental similarities.

Contributing to the unpredictability of imitative words is the fact that they are sometimes widely borrowed among neighboring languages, even when ordinary cultural borrowings do not take place. Baby-words, including those for *father* and *mother* (like *pa*, *papa*, *dad* and *ma*, *mama*), agree with imitatives in this tendency.

4.8. Although words were from the first generally avoided if their usage was too closely bound up with peculiarities of language morphology, a few were included: *at*, *in*, *because if*, *and*. These are all eliminated in the revision.

5. If the lexicostatistic test list is to be valid for different magnitudes of time depth, it has to be made up of elements having approximately the same likelihood to persist. To the extent that this condition is not met, there will be a tendency for the less persistent items to be lost early, thereby leaving a residual list of gradually increasing durability until the retention rate reaches a maximum definitely above the initial rate. Lees' pilot study tells little about this problem, because the control cases are all of relatively short duration, from 1.0 to 2.2 thousands of years of unilinear change. The application of lexicostatistic datings is mainly to cases of divergence in a pair of languages, rather than of change in a single line, and the time depths have extended up to about 8000 years. Now, a millenium of unilinear change is equivalent to five centuries of divergence, and 2.2 millenia in one line are equivalent to 1.1 of divergence. Therefore, the control studies, being limited to the kind of cases historically known, cover only a small fraction of the required range.

In order to overcome the lack of long-range control cases, one may measure the

persistence rates of individual test items. The procedure is suggested in the following:

Old	3	4	7	18	22	33	26	23	37	42	(Total 215)
New	0	0	1	6	8	8	14	12	23	20	(Total 92)

"The stability of the items . . . needs to be objectively tested by noting how often and for how long they are retained in a number of historically known situations. A stability score for individual items could be calculated, and this score taken into account in constructing the improved test list". (Lexico-Statistic Dating, 457). With respect to the number of languages which should be included in the test, it is still impossible to carry out this procedure. But it is possible to measure item persistence in the control material, and we have done so by scoring decades of retention over total decades observed in the control languages (see the 'Persistence' column in Table 2). We have limited the count to eight languages in order to avoid overlapping cases (explained farther on). We have used Lees' time spans even though in the case of Coptic there is reason to believe that a longer separation time is involved.

The value of the persistence score is very provisional, because of the small number of languages. Thus, although there is no such thing as an immutable vocabulary item, we have a number of words which score 100%. Of course this situation would be changed if we had a sufficiently large number of control cases. It is interesting that the numerals comprise a large part of the perfect scores, since experience with linguistic stocks in other parts of the world, involving economies which are either more primitive or which have been developed in relatively recent times, demonstrate considerable instability in their numerals.

The way in which we can use these tentative persistence scores is by comparing their distribution in the old and new lists. The result, with the scores being grouped by decades (1–10%, 11–20%, etc.), is as follows:

The persistence scores of the old list are seen to have two areas of concentration, 51–60% and 81–100%, and to be widely spread over the whole possible range of values. The indications are of mixed criteria and of heterogeneous persistence values. If the new list had been made artificially, by merely cutting out words in such a way as to give a good distribution, the apparently improved distribution would be meaningless and there could be no expectation that its lexicostatistic application would give more reliable results than the first. However, not one of the words was dropped by virtue of its individual score. Instead the changes were made by categories and according to general criteria, as discussed above, deriving from definite methodological needs, all of which had made themselves felt in the course of lexicostatistic work long before the persistence scores were taken. Under these circumstances, the new list seems to offer considerable promise for rectifying much if not all of the skewing of the old one.

It will be seen in the further discussion that the new test list has a higher retention rate than the old, approximately 86%, in place of approximately 81% per thousand years. This is in keeping with the item persistence scores we have just observed, and it gives us a means of measuring the suspected skewing effect in the old list. Thus, if approximately half the old list has a retention index of 86% and the whole list 81%, then the remaining half of the list evidently has an average rate of 76%. Translated into divergences instead of unilinear rates, this gives us 74% for the upper half, 58% for the lower half, and 66% for the whole list. We can project these figures into larger time spans and thereby learn the

extent of skewing, calculated very approximately as follows:

Centuries	10	20	30	40	50	60	70	80
Upper half	74	55	40	30	22	16	12	9
Lower half	58	34	20	11	6.6	3.8	2.2	1.3
Complete list	66	45	30	21	14	10	7	5
Old calculation	66	44	28	19	12.5	8.3	5.6	3.6
Underestimation in centuries	0	1	2	3	4	6	8	12

Thus, disregarding all other deficiencies and examining only the skewing effect due to mixed persistence rates in the old list, we see that using a uniform cognate retention rate of 66%, instead of recognizing the mixed persistence in the old items gives us a definite underestimation which runs from a century in 20 to 3 in 20, as one increases the time depth to the limits of the lexicostatistic method. This measure does not, however, include all of the skewing but only the difference between the old calculation and the adjustment indicated by a rough persistence analysis.

We can be sure that the new list still gives a certain amount of skewing, for it is not to be imagined, nor is it suggested by the persistence scores, that its component items are of uniform stability. We can, however, deduce that the new skewing is considerably less than the old. Also, since the range of variation is considerably reduced in the new list and since it is better concentrated around a single norm, we can infer that the residual skewing may be a small fraction of the old. Be this as it may, it is pointless to attempt to calculate (presumably by differential calculus) the residual effect, until we have truly representative persistence scores based on a far larger number of control cases. Once we have an accurate measure of the skewing, the effect could be incorporated into the time depth equation or, more conveniently, one could use prepared tables showing time and expected cognate ratio.

6. Some aspects of the procedure in the control study need revision.

6.1. Some of the control cases are related languages which have shared a part of their history. For example, Lees' case 2 is "Plautine Latin of 200 B.C.: Early Modern Spanish of 1600 A.D." Case 3 is "Plautine Latin: Molière's French of 1650 A.D." These cases were the same for about 1000 years, and only after this common period did they develop in relative independence. To use both is therefore in large part a duplication of a single instance, and there is of course no justification for counting some of the examples more than once. To avoid this, one could break up the time spans into three non-overlapping segments: (a) Plautine Latin to Late Latin, (b) Late Latin to early modern Spanish, (c) Late Latin to Molière French. Or, alternatively, into two: (a) Plautine Latin to early modern Spanish, (b) Late Latin to Molière French. And similar treatment would be required for Portuguese, Catalan, Italian, and the two types of Greek. This procedure would require new test lists for at least one suitable stage of Late Latin, for Iberian Romance, and for an intermediate stage of Greek suitable for separating Athenian from Cypriote. In the absence of such test lists, we must eliminate all but one of each set of interlocking cases.

6.2. The control cases are intended to be single languages taken in their earlier and later stages, and for the most part the cases can be checked historically to give adequate or nearly adequate assurance that they meet the requirement. In the Latin-French sequence, for example, the language is known from written records throughout the intervening period, so that we can be fairly sure we are dealing essentially with a

single line of development. But, in the case of Middle Egyptian and Coptic, there is a long break in the record between 1700 B.C. and 300 A.D. We cannot be sure that Coptic is a lineal descendant of Middle Egyptian. It may derive from a speech form which in 1700 B.C. was already divergent from the Egyptian dialect preserved in the written records. This uncertainty may be important in evaluating our results, for we find that the percentage of common elements in Egyptian-Coptic is much lower than any of the other examples would lead us to expect. In fact, if we evaluate the cognate percentage in the light of the average retention rate derived from other examples, we find that Egyptian and pre-Coptic of 1700 B.C. must already have been divergent at that time by the equivalent of ten centuries of independent development. It is possible that a similar situation applies, though to a smaller degree, in some of the other cases. For the purposes of the present study, we eliminate Coptic.

6.3. In Lees' study, each control case is given equal weight regardless of the length of time covered. This is perhaps acceptable as a simplification of the mathematics, in view of the fact that the results may come out near each other, but more accurate figuring requires weighting each case according to the period of time. The present study follows the latter procedure.

6.4. The dating of language samples may present difficulties, and Lees recognizes that there may be dating errors in his study. However, differences of several decades to a century in individual examples do not actually have any very appreciable effect on the final figures and in any event the present study has no basis for departing in any way from Lees' dates.

6.5. There is a certain danger of subjectivism in filling out the test lists for given languages and in scoring cognates between a pair of lists. Arndt, making independent calculations for the Germanic languages, concluded with reference to the earlier scores that there "seemed to be a consistent inclination . . . to find more cognates than a strict analysis permits." His rate for Swedish is 79.7% instead of 85.4%, for German 82.0% in place of 85.4%, for English 67.8% or 74.8% instead of 76.6%. Since Arndt was able to give more time and care to the matter than Lees and his collaborators, and since he approached his task in a most critical manner, it is reasonable to suppose that his figures are more accurate than the old. However, since his work sheets were unavailable during the preparation of the present paper, it has been necessary to use the old scoring. This is unfortunate since it is precisely in connection with Swedish and German that the largest deviations are found. Arndt's scoring of the Germanic languages would to some extent narrow the dispersion in Lees' study, and if a similar difference holds in the new diagnostic list, would make a dramatic improvement in the new dispersion.

Another control case which should be checked is Chinese, based on lists kindly supplied by C. Y. Fang and scored by the present author, but not re-checked in the careful way that is necessary to guarantee a maximally dependable calculation.

6.6. The problems discussed thus far are concerned with accuracy of supplying the diagnostic words and the procedure for deriving the index. We must now touch on the influence of possible factors other than time itself. Whatever modifying factors there may be, they will have to be discovered and analyzed in their quantitative and qualitative impact, in order to find a technique either for accurately determining their effect on the retention rate or for otherwise subtracting their influence from the net result. In large part the latter has been done already by omitting culture words and certain other categories from the diagnostic list, but there is no proof that this

eliminates all side factors. Arndt reports having established a modification in the English retention rate resulting from interdialect borrowing. He gives the gross retention rate as 67.8% per 1000 years and the rectified rate, allowing for the interdialect borrowings, as 74.8%. Of all the Germanic languages which he studied, he mentions this special influence only for English.

7. We may now compare the old and new tests, to determine how much improvement has been achieved. We note that the new list contains more universal and more easily defined elements, which may be more easily supplied in all languages. Also, by narrowing the persistence range of the component items, it reduces much of the skewing inherent in the old list for long divergence periods. In large part these advantages would not show in a statistical analysis based on our seven control languages, because the languages belong to relatively uniform culture areas and stages and because their time spans are all relatively short. Other advantages, such as those due to greater freedom from unpredictable culture terms and the like, should be reflected to some extent in a comparison of average rates and deviations for the two lists. On the other hand, the fewness of the control cases and the uncertainties as to scoring, may obscure the results.

The results for the two diagnostic lists—cognate counts, persistence rates per 1000 years for each language and for the languages taken together, and the standard deviation—are as follows:

Language	Millenia	Old score		New score		Old rate	New rate
		count	%	count	%	%	%
Swedish	1.02	176/207	(85.0)	83/88	(94.3)	85.4	94.4
German	1.1	180/214	84.2	81/89	89.0	85.4	90.0
English	1.0	160/209	76.6	75/87	86.2	76.6	86.2
Rumanian	2.15	117/209	56.0	63/88	71.6	76.4	85.6
French	1.85	125/200	62.5	66/89	74.2	77.6	85.1
Athenian	2.07	147/213	69.0	64/90	71.1	83.6	84.8
Chinese	1.0	167/210	79.6	74/91	81.5	79.6	81.5
For seven cases (totalling 10.19 millenia)							
Average rate						80.2	86.4
Standard deviation						3.65	3.33
Omitting Swedish (6 languages over a total of 9.17 millenia)							
Average rate						79.7	85.4
Standard deviation						3.44	2.09
For three 2000-year cases (total millenia 6.07)							
Average rate						79.0	85.4
Standard deviation						3.19	0.4

One of the evident effects of the revised list is to raise the index of retention by about 6%. This in itself is not necessarily an advantage. A higher index is helpful with respect to extending the limits of the test to more remote time depths. A lower index, if equally accurate, would allow for finer distinctions in time span within its usable range. Abstractly considered, it might be desirable to have a succession of different tests with higher and lower retention rates to meet different possible needs, but suitable diagnostic items are not indefinitely available and it probably would be difficult to set up more than say two differentiated time scales. If this were done, the two lists would presumably have to avoid overlapping. In other words, our old and new lists could not be considered as forming a battery of parallel tests, to be employed according to the time depths to be measured. Instead, our problem is to decide which of the two is the more suitable and accurate as a test with a relatively high retention rate.

The measure of accuracy we have used here is the standard deviation, defined as the square root of the average square of the deviations from the mean. (This is not synonymous with standard error, the measure Lees used in his study, which is by definition a minor fraction of the standard deviation.) The rule is that the smaller the deviation the more dependable the test.

These figures show that the deviations of the old retention rates are somewhat scattered within their maximum limits. Even Coptic, with the low apparent retention of 74.9% (correcting a small arithmetic error in Lees' paper, pointed out by Louis Cote and Albert Reid), which we believe may be due to hidden divergence, would accommodate to the other figures without greatly increasing the standard deviation. Eliminating Swedish, which we have reason to believe has been scored too high by about 5 percentage points, gives only a moderate reduction in the deviation. And eliminating the other three cases suspected of inaccurate scoring gives again only a small further improvement.

The revised test sharply separates the inaccurate scores from the others. Rumanian, French and Athenian are very closely clustered. It is interesting that these are the three cases involving time periods in the neighborhood of 2000 years. The circumstance may be partly connected with the fact that errors in short-term scores may be cancelled out over a longer period, or it may be due to better scoring or to combination of the two factors.

If we included Coptic, the new test would give a deviation a little under that of the old test, close to 4%. The omission of Swedish along with Coptic cuts the deviation almost in half. And the further elimination of the remaining three cases whose scores are somewhat suspect, brings the standard deviation down to 0.4%. Too much importance cannot be attached to this encouragingly low deviation until much more extensive tests are carried out. If this is anywhere near a correct evaluation of the revised test, lexicostatistics becomes almost a precision instrument. In view of the fact that the revised test has great advantages over the old in adaptability of the test items, the revision would be distinctly worthwhile even if it did not markedly increase the accuracy. But there is no reason to doubt that definite improvement in statistical accuracy goes with the other improvements.

At the same time, we are encouraged to seek still further improvements, and above all to hope for the major research which can definitely establish the capabilities and the limitations of the lexicostatistic method.

TABLE 1—Item persistence by semantic groups

a) Personal Pronouns

I	100%
thou	92
we	92
he	70
ye	67
they	50

b) Interrogatives

who	83%
where	75
what	74
when	74
how	70

c) Correlatives

and	57%
if	31
because	25

d) Locatives

at	100%
in	57
with	51

e) Location

there	88%
far	74
near	51
right (side)	51
here	49
that	42
this	34
left (side)	33

f) Position and Movement

come	100%
sit	100
give	85
fly	82
stand	77
hold	69
fall	67
swim	59
turn	59
walk	59
throw	49
pull	43
float	40
flow	33
lie	33
push	17

g) Manipulations

wash	83%
split	75
tie	74
hit	69
wipe	58
cut	56
rub	50
dig	42
scratch	31
squeeze	28

h) Time Periods

year	100%
day	85
night	82

i) Numerals

one to ten, incl.	100%
twenty	83
hundred	82

j) Quantitatives

all	92%
few	74
many	58
some	40

k) Size

wide	85%
thick	83
long	82
thin	79
narrow	59
big	48
small	41
short	25

l) Natural Objects and Phenomena

ice	100%
salt	100
star	100
sun	100
wind	100
sky	92
cloud	83
rain	83
water	83
sea	82
smoke	81
snow	79
sand	68
stone	68
mountain	67
ashes	66
earth	66
dust	56
lake	56
fog	51
river	51
fire	33

m) Plants & Plant Parts

bark	100%
leaf	100
grass	83
tree	83
root	82
flower	75
woods	68
seed	66
berry (fruit)	59
stick	27

n) Animals

worm	100%
snake	91
louse	80
fish	66
dog	50
animal	42
bird	40

o) Persons

person (human being)	74%
woman	71
child	50
man	45

p) Body Parts and Substances

blood	100%
ear	100
hand	100
tongue	100
tooth	100
foot	90
egg	84
back	83
tail	83
meat (flesh)	77
eye	74
feather	74
skin	74
bone	71
head	71
mouth	68
nose	66
wing	66
heart	65
fat	56
guts	55
belly	50
neck	49
hair	48
liver	40
leg	17

q) Body Sensations and Activities

drink	92%
die	92
hear	85
see	82
sleep	82
live	75
eat	68
know	59
bite	57
fear	43
think	33
breathe	25
vomit	19
smell	0

r) Oral Activities

laugh	92%
sing	83
suck	75
cry	68
spit	56
speak	56

s) Colors

black	83%
green	83
red	66
white	51
yellow	51

t) Descriptives

old	100%
dry	89
good	83
new	82
warm	79
rotten	66
cold	65
sharp	58
right (correct)	50
straight	49
smooth	41
bad	40
wet	33
dull	20
dirty	10

u) Kinship

brother	100%
sister	100
father	83
mother	83
husband	57
wife	51

v) Cultural Objects and Activities

sew	100%
rope	85
shoot	60
hunt	59
cook	58
count	51
play	50
clothing	42
work	37
dance	33
spear	33
stab	30
fight	9

w) Miscellaneous

name	100%
other	92
not	83
burn	67
blow	66
freeze	65
swell	51
road	33
kill	25

TABLE 2—Persistence of diagnostic items in Coptic, Chinese, English, German, Swedish, French, Rumanian and Athenian; retention or loss in Spanish, Portuguese, Catalan, Italian and Cypriote Greek

	Cop	Chi	Eng	Ger	Swe	Fre	Rum	Ath	Persistence		Spa	Por	Cat	Ita	Cyp
									Decades	Pct					
*all	+	−	+	+	+	+	+	+	1139/1239	92	+	+	+	+	+
and	−	−	+	+	+	+	−	+	704/1239	57	+	+	+	+	+
animal	−	+	−	+	+	−	−	+	519/1239	42	+	0	0	+	+
*ashes	−	+	+	+	+	+	+	−	812/1239	66	+	+	+	+	−
at	+	+	+	+	0	0	0	−	837/837	100	0	+	+	+	−
back	+	+	+	+	+	+	+	−	1032/1239	83	−	−	−	−	−
bad	−	+	−	−	−	+	−	+	492/1239	40	+	+	−	−	+
*bark	0	+	0	+	0	+	+	+	817/817	100	+	+	+	+	−
because	−	+	−	−	−	−	−	+	307/1239	25	−	−	−	−	+
*belly	+	−	−	−	−	+	−	+	612/1239	50	+	+	+	+	+
berry	0	+	+	+	+	+	−	−	597/1019	59	+	+	−	+	−
*big	−	+	−	−	+	+	−	+	594/1239	48	+	+	+	+	+
*bird	−	+	−	+	+	+	−	−	497/1239	40	−	+	+	+	−
*bite	−	−	+	+	+	+	0	+	704/1024	57	+	+	0	+	0
*black	+	+	+	+	+	+	+	−	1032/1239	83	+	−	+	+	−
*blood	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
blow	−	+	+	+	+	+	+	−	812/1239	66	+	+	−	+	−
*bone	+	+	+	−	+	+	+	−	922/1239	71	+	+	+	+	−
breathe	−	+	−	+	+	−	−	−	312/1239	25	+	+	+	+	−
brother	+	+	+	+	+	+	+	+	1239/1239	100	−	−	−	+	0
*burn	−	+	+	+	+	−	+	+	834/1239	67	+	+	+	−	+
child	−	+	+	+	+	−	−	+	619/1239	50	−	−	−	−	+
clothing	+	+	+	−	+	−	−	−	522/1239	42	−	+	+	+	−
*cloud	+	+	−	+	−	+	+	+	1037/1239	83	+	+	+	+	+
*cold	−	+	+	+	+	+	−	+	804/1239	65	+	+	+	+	+
*come	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
cook	+	+	+	+	−	+	−	−	715/1239	58	+	−	−	−	−
count	+	+	−	+	−	−	−	+	637/1239	51	−	−	−	−	+
cry	+	+	+	+	+	−	−	+	839/1239	68	−	−	+	−	+
cut	+	0	−	+	+	−	−	+	639/1139	56	−	−	−	−	+
dance	−	+	−	−	+	−	−	+	409/1239	33	−	−	−	−	+
day	+	+	+	+	+	−	+	+	1054/1239	85	+	+	+	−	+
*die	+	+	−	+	+	+	+	+	1139/1239	92	+	+	+	+	+
dig	−	+	−	+	+	−	−	+	519/1239	42	−	−	−	−	+
dirty	0	+	−	−	−	−	−	−	100/1019	10	−	−	−	−	−
*dog	−	−	−	+	+	+	+	−	612/1239	50	−	+	−	+	−
*drink	+	−	+	+	+	+	+	+	1139/1239	92	+	+	+	+	+
*dry	+	+	+	−	+	+	0	+	914/1024	89	+	+	+	+	+
dull	0	+	−	−	+	−	−	−	202/1019	20	−	−	−	−	−
dust	−	+	+	+	−	+	−	+	702/1239	56	+	+	+	+	+
*ear	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*earth	+	+	+	+	+	+	−	−	817/1239	66	+	+	+	+	−
*eat	+	+	+	+	+	−	−	+	839/1239	68	+	−	−	−	+
*egg	+	−	−	+	+	+	+	+	1039/1239	84	+	+	+	+	+
eight	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+

TABLE 2—Persistence of diagnostic items—*Continued*

	Cop	Chi	Eng	Ger	Swe	Fre	Rum	Ath	Persistence		Spa	Por	Cat	Ita	Cyp
									Decades	Pct					
*eye	−	−	+	+	+	+	+	+	919/1239	74	+	+	+	+	+
fall	−	+	+	+	+	−	+	+	834/1239	67	+	+	+	+	+
far	+	+	+	−	+	+	−	+	914/1239	74	−	+	+	+	+
*fat(grease)	−	+	+	−	+	+	−	+	694/1239	56	−	−	−	−	+
father	+	+	+	+	+	+	−	+	1024/1239	83	+	+	+	+	+
fear	−	−	−	+	−	−	+	+	532/1239	43	+	+	+	+	+
*feather	−	−	+	+	+	+	+	+	919/1239	74	+	+	+	+	+
few	+	+	+	−	+	+	−	+	914/1239	74	+	+	+	+	+
fight	−	0	+	−	−	−	−	−	100/1139	9	−	−	−	−	−
*fire	−	+	+	+	+	−	−	−	412/1239	33	−	−	−	−	−
*fish	−	+	+	+	+	+	+	−	812/1239	66	+	+	+	+	−
five	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
float	0	−	+	−	+	−	−	+	409/1019	40	+	−	−	−	+
flow	−	+	+	+	+	−	−	−	412/1239	33	−	−	−	−	−
flower	+	+	−	+	+	+	+	−	932/1239	82	+	+	+	+	−
*fly	−	+	+	+	+	+	+	+	1019/1239	82	+	+	+	+	+
fog	0	+	+	+	−	−	−	+	517/1019	51	+	+	−	+	+
*foot	+	−	+	+	+	+	0	+	924/1024	90	+	+	+	+	+
four	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
freeze	−	+	+	+	+	+	−	+	804/1239	65	+	+	+	+	+
*give	+	+	+	+	+	−	+	+	1054/1239	85	+	+	+	+	+
*good	+	+	+	+	+	+	+	−	1032/1239	83	+	+	+	+	−
grass	+	+	+	+	+	+	+	−	1032/1239	83	+	+	+	+	−
*green	+	+	+	+	+	+	+	−	1032/1239	83	+	+	+	+	−
guts	+	+	0	−	+	−	−	+	629/1139	55	−	−	−	−	+
*hair	−	+	+	+	+	+	−	−	597/1239	48	−	+	+	+	−
*hand	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
he	+	−	+	+	+	0	−	+	739/1054	70	0	−	−	−	+
*head	0	+	+	−	+	−	+	+	724/1019	71	+	+	+	−	+
*hear	+	+	+	+	+	−	+	+	1054/1239	85	+	+	−	−	+
*heart	−	+	+	+	+	+	−	+	804/1239	65	+	+	+	+	+
here	−	−	+	+	+	+	0	+	497/1024	49	−	0	0	0	−
hit	+	−	−	+	+	−	+	+	854/1239	69	−	−	−	−	+
hold	−	−	+	+	+	+	+	0	712/1032	69	+	−	−	+	0
how	−	−	+	+	+	0	+	+	734/1054	70	0	+	+	+	+
hundred	+	+	+	+	0	+	−	+	922/1137	82	+	+	+	+	+
hunt	−	+	+	+	−	−	+	+	732/1239	59	−	−	−	−	+
husband	+	+	−	−	−	+	−	+	712/1239	57	+	+	+	+	+
*I	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
ice	0	+	+	+	+	+	+	+	1019/1019	100	−	−	+	+	+
if	−	+	+	−	−	+	−	−	385/1239	31	+	+	+	+	−
in	−	−	+	+	+	+	+	−	712/1239	57	+	+	−	+	−
*kill	−	+	−	+	+	−	−	−	312/1239	25	−	−	−	+	−
*know	−	+	+	+	+	0	+	−	627/1054	59	0	−	−	0	−
lake	−	+	−	+	0	−	+	+	632/1137	56	+	+	+	+	+

TABLE 2—Persistence of diagnostic items—*Continued*

	Cop	Chi	Eng	Ger	Swe	Fre	Rum	Ath	Persistence		Spa	Por	Cat	Ita	Cyp
									Decades	Pct					
laugh	+	+	+	+	−	+	+	+	1137/1239	92	+	+	+	+	+
*leaf	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
left(side)	−	+	−	−	+	−	−	+	409/1239	33	−	−	−	+	+
leg	−	+	−	+	−	−	−	−	210/1239	17	−	−	−	−	−
*lie	−	+	+	+	+	−	−	−	412/1239	33	−	0	+	+	−
live	+	−	+	+	+	+	−	+	924/1239	75	+	+	+	+	+
*liver	0	+	+	+	+	−	−	−	412/1019	40	−	−	−	−	−
*long	−	+	+	+	+	+	+	+	1019/1239	82	−	+	−	+	+
*louse	0	+	+	+	+	+	+	−	812/1019	80	+	+	+	+	+
*man	−	+	+	+	0	−	−	+	517/1137	45	−	−	−	−	+
*many	−	+	+	−	+	−	+	+	724/1239	58	+	+	+	+	+
*meat(flesh)	+	+	−	+	+	−	+	+	954/1239	77	+	+	+	+	+
mother	+	+	+	+	+	+	−	+	1024/1239	83	+	+	+	+	+
*mountain	+	+	−	+	−	+	+	−	830/1239	67	+	+	+	+	+
*mouth	+	+	+	+	+	−	−	+	839/1239	68	−	−	−	−	+
*name	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
narrow	−	−	+	+	+	−	+	+	734/1239	59	−	−	−	−	+
near	−	+	+	+	+	−	+	−	627/1239	51	−	−	−	−	−
*neck	−	−	−	+	+	+	−	+	604/1239	49	+	+	+	+	+
*new	−	+	+	+	+	+	+	+	1019/1239	82	+	+	+	+	+
*night	−	+	+	+	+	+	+	+	1019/1239	82	+	+	+	+	+
nine	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*nose	−	+	+	+	+	+	+	−	812/1239	66	+	+	+	+	−
*not	+	+	+	+	+	+	+	−	1032/1239	83	+	+	+	+	−
old	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*one	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
other	+	−	+	+	+	+	+	+	1139/1239	92	+	+	+	+	+
*person(human being)	−	+	0	+	0	0	+	+	632/852	74	0	+	+	+	+
play	−	+	+	+	+	−	−	+	619/1239	50	−	−	−	−	+
pull	−	+	−	+	+	−	+	−	527/1239	43	−	−	−	−	−
push	−	+	−	+	−	−	−	−	210/1239	17	+	0	−	−	−
*rain	+	+	+	+	+	+	+	−	1032/1239	83	+	+	+	+	−
*red	+	+	+	+	+	+	−	−	817/1239	66	−	−	−	−	−
right(side)	−	+	+	+	+	−	+	−	629/1239	51	0	−	−	+	−
right(correct)	+	+	−	−	+	−	−	+	617 1239	50	−	−	−	+	+
river	+	+	−	+	−	−	−	+	637/1239	51	−	−	−	+	+
*road(path)	−	+	+	+	+	−	−	−	412/1239	33	−	−	−	−	−
*root	−	+	+	+	+	+	+	+	1019/1239	82	+	+	+	+	+
rope	+	+	+	+	+	−	+	+	1054/1239	85	−	−	−	+	+
rotten	−	0	−	+	+	0	+	+	634/954	66	−	+	+	−	+
rub	−	−	−	+	−	0	+	+	532/1054	50	0	+	+	+	+
salt	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*sand	+	+	+	+	+	−	−	+	839/1239	68	+	+	−	+	+
scratch	−	0	0	+	−	−	−	+	317/1039	31	−	−	−	−	+
sea	−	+	+	+	+	+	+	+	1019/1239	82	+	+	+	+	+

TABLE 2—Persistence of diagnostic items—*Continued*

	Cop	Chi	Eng	Ger	Swe	Fre	Rum	Ath	Persistence		Spa	Por	Cat	Ita	Cyp
									Decades	Pct					
*see	−	+	+	+	+	+	+	+	1019/1239	82	+	+	+	+	+
*seed	−	+	+	+	+	+	+	−	812/1239	66	+	+	−	+	−
seven	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
sew	0	+	+	+	0	+	+	+	917/917	100	+	+	+	+	+
sharp	+	−	+	+	+	+	−	−	717/1239	58	+	−	−	+	−
shoot	+	+	+	+	+	0	−	−	632/1054	60	0	−	−	−	−
short	−	+	+	+	−	−	−	−	310/1239	25	−	−	−	−	−
sing	+	+	+	+	+	+	+	−	1032/1239	83	+	+	+	+	−
sister	+	+	+	+	+	+	+	+	1239/1239	100	−	−	−	+	0
*sit	+	+	+	+	+	+	+	+	1239/1239	100	−	+	+	+	+
six	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*skin	−	+	−	+	+	+	+	+	919/1239	74	+	+	+	+	+
sky	+	+	−	+	+	+	+	+	1139/1239	92	+	+	+	+	+
*sleep	−	+	+	+	+	+	+	+	1019/1239	82	+	+	+	+	+
*small	−	+	+	−	+	−	−	+	509/1239	41	−	−	−	−	+
smell	−	−	−	−	−	−	−	−	0/1239	0	+	−	0	−	−
*smoke	−	+	0	+	+	+	+	+	919/1139	81	+	+	+	+	+
smooth	−	+	+	−	+	−	−	+	509/1239	41	−	−	−	−	−
snake	+	+	+	−	+	+	+	+	1129/1239	91	+	−	+	+	−
snow	0	+	+	+	+	+	−	+	804/1019	79	+	+	+	+	+
some	+	−	+	−	+	0	−	−	422/1054	40	0	+	+	+	−
speak	+	+	+	+	+	−	−	−	632/1239	51	+	−	−	−	−
spear	−	+	+	+	+	−	−	−	412/1239	33	−	−	−	−	−
spit	−	+	+	+	−	−	+	+	732/1239	56	−	+	+	+	+
split	+	+	−	+	+	+	−	+	924/1239	75	+	−	−	−	+
squeeze	−	+	−	0	−	−	+	−	315/1129	28	−	−	+	+	−
stab	−	+	−	+	+	0	−	−	312/1054	30	0	−	−	−	−
*stand	+	−	+	+	+	−	+	+	954/1239	77	−	+	+	+	+
*star	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
stick	−	+	0	+	+	−	−	−	412/1139	27	+	+	+	−	−
*stone	+	+	+	+	+	−	−	+	839/1239	68	−	−	−	−	−
straight	−	+	−	−	+	+	+	−	602/1239	49	+	+	+	+	−
suck	+	−	+	+	+	+	+	−	932/1239	75	−	−	0	+	−
*sun	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
swell	+	+	+	+	+	−	−	−	632/1239	51	−	−	−	−	−
*swim	−	−	+	+	+	−	+	+	734/1239	59	+	+	+	+	+
*tail	+	+	+	−	−	+	+	+	1027/1239	83	+	−	+	+	+
ten	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*that	−	−	+	+	+	−	−	+	519/1239	42	+	0	0	0	+
there	+	−	+	+	+	0	0	+	739/839	88	0	+	+	+	+
they	0	−	−	+	+	0	−	+	419/834	50	0	−	−	−	+
thick	+	+	+	+	+	+	−	+	1024/1239	83	+	+	0	0	+
thin	0	+	+	+	+	+	−	+	804/1019	79	−	−	−	−	+
think	−	+	+	−	−	−	−	+	407/1239	33	−	−	−	−	+
*this	−	−	+	+	−	−	−	+	417/1239	34	−	−	0	−	+

TABLE 2—Persistence of diagnostic items—*Continued*

	Cop	Chi	Eng	Ger	Swe	Fre	Rum	Ath	Persistence		Spa	Por	Cat	Ita	Cyp
									Decades	Pct					
*thou	+	−	+	+	+	+	+	+	1139/1239	92	+	+	+	+	+
three	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
throw	−	−	−	+	+	+	−	+	604/1239	49	+	+	−	+	+
tie	−	−	+	+	+	+	+	+	919/1239	74	−	−	+	+	+
*tongue	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*tooth	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*tree	−	+	+	+	+	+	−	+	1024/1239	83	+	+	+	+	+
turn	−	+	−	+	+	−	+	+	734/1239	59	−	−	−	−	+
twenty	+	+	+	+	+	+	−	+	1024/1239	83	+	+	+	+	+
*two	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
vomit	−	+	−	−	+	0	−	−	202/1054	19	0	+	−	+	−
*walk	−	+	−	+	+	−	+	+	734/1239	59	+	−	−	−	+
*warm(hot)	−	+	+	+	+	+	+	0	812/1032	79	+	−	+	+	+
wash	+	+	+	+	+	+	−	+	1024/1239	83	+	+	−	+	+
*water	+	+	+	+	+	+	+	−	1032/1239	83	+	+	+	+	−
*we	+	−	+	+	+	+	+	+	1139/1239	92	+	+	+	+	+
wet	−	+	+	+	+	−	−	−	412/1239	33	−	−	−	−	−
*what	−	−	+	+	+	+	+	+	919/1239	74	+	+	+	+	+
when	−	−	+	+	+	+	+	+	919/1239	74	+	+	+	+	+
where	+	−	+	+	+	+	−	+	924/1239	75	−	−	−	0	+
*white	−	+	+	+	+	−	+	−	627/1239	51	−	−	−	−	−
*who	+	+	+	+	+	+	+	−	1032/1239	83	+	+	+	+	−
wide	+	+	+	+	+	−	+	+	1054/1239	85	−	−	−	−	+
wife	+	+	+	−	−	−	−	+	627/1239	51	−	−	−	−	+
wind	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
wing	+	+	−	−	+	+	−	+	814/1239	66	+	+	+	+	+
wipe	0	−	+	+	0	0	+	−	425/732	58	0	−	−	−	−
with	−	−	−	+	+	−	+	+	634/1239	51	+	+	−	+	+
*woman	+	+	0	−	+	+	−	+	814/1139	71	−	−	−	−	+
woods	+	+	+	+	+	−	−	+	839/1239	68	−	−	−	−	+
work	−	0	+	+	−	−	−	+	417/1139	37	−	−	−	−	+
worm	+	+	+	+	+	+	+	+	1239/1239	100	−	+	−	+	+
ye	+	−	+	+	−	+	+	−	830/1239	67	+	+	+	+	−
year	+	+	+	+	+	+	+	+	1239/1239	100	+	+	+	+	+
*yellow	+	+	+	+	+	−	−	−	632/1239	51	−	−	−	−	−

LEXICO-STATISTICAL SKEWING FROM DIALECT BORROWING

SARAH C. GUDSCHINSKY
SUMMER INSTITUTE OF LINGUISTICS

0. Introduction
1. The dip
2. Lexical dialects
3. Phonological dialects
4. Discrepancy between lexical and phonological dialects
5. Conclusion

0. Introduction. It is the purpose of this paper: (a) to propose a technique for discussing degrees of lexical relationship between dialects in terms of DIPS,[1] which, although based on lexico-statistical techniques, provide an objective measure of relationships free from inherent implications of absolute time (though not of relative time); (b) to show, for six closely related Mazatec[2] dialects, successive periods of lexical development as deduced from lexico-statistical data in terms of dips; (c) to show for the same dialects successive periods of phonological development as deduced from data obtained by traditional comparative methods; (d) to point out the discrepancy between the order of the implied periods of lexical development and the phonological development, and to discuss this discrepancy as caused by a skewing of the lexical data due to recent dialect borrowing; and (e) to discuss the theoretical implications of this lexical skewing with reference to the validity of the lexico-statistical method.

The dialects included in this study are those spoken in the villages of Huautla de Jiménez, San Mateo Huautla, San Miguel Huautla, San Miguel Soyaltepec, San Pedro Ixcatlán, and San Felipe Jalapa de Díaz.[3] The approximate geographical location of these villages is shown on Map A. Huautla, San Mateo, and San Miguel are mountain villages located within two to four hours walking distance from each other. Soyaltepec, Ixcatlán, and Jalapa are lowland villages, located more than a day's walk from the mountain villages, and separated from each other by three to six hours walking distance.

An initial lexico-statistical study was made, comparing each of the six dialects with each of the others—a total of fifteen comparisons.[4] The pairs of dialects are listed in Table 1 according to the percentages of apparent cognates they have in common, so that the higher on the scale a particular pair of dialects is listed, the fewer cognates the dialects have in common. Since, within the basic assumptions of lexico-statistics, the lower the percentage of cognates a pair of dialects have in common,

[1] I am indebted to Kenneth L. Pike specifically for the suggestion which led to the development of 'dips', and also for other suggestions and criticisms which made this paper in its present form possible.

[2] Mazatec is a language of Southern Mexico spoken by an estimated 85,000 Indians of the State of Oaxaca.

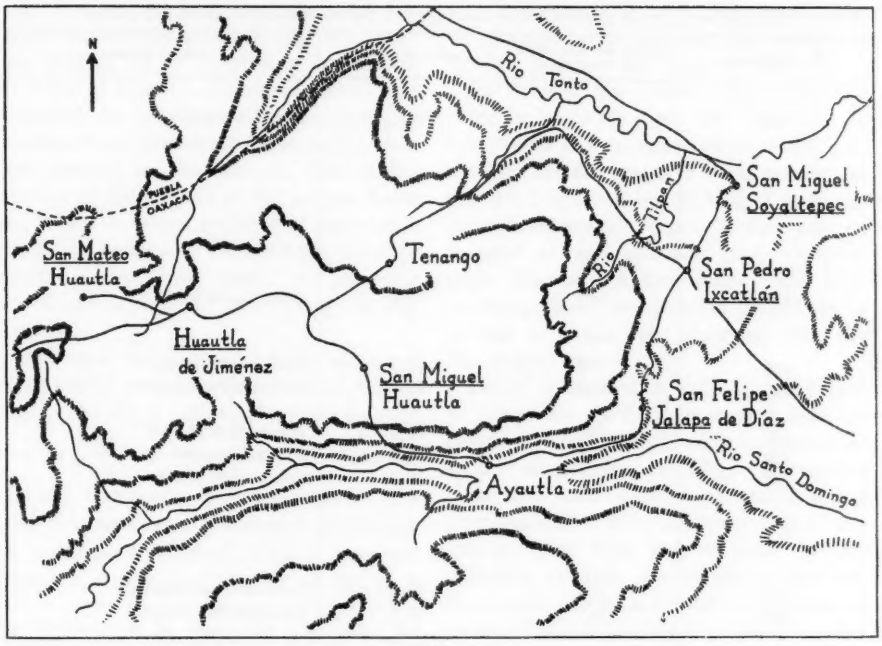
[3] The data were gathered during field trips under the auspices of the Summer Institute of Linguistics. The Huautla data were gathered in 1948–54. The phonemic analysis is that of Kenneth L. Pike and Eunice V. Pike, Immediate Constituents of Mazateco Syllables, IJAL 13.78–91 (1947).

The San Miguel data were gathered in 1950–51, the San Mateo data in 1951–52, and the Soyaltepec data in 1954, in collaboration with Eunice V. Pike.

The Ixcatlán data, and some supplementary Soyaltepec data, are from the field notes of George Cowan.

The Jalapa data were prepared by Paul Kirk, who also has made a tentative phonemic analysis of the Jalapa dialect.

[4] Swadesh's 200 word list was used for these comparisons. See Morris Swadesh, The Language of the Archaeologic Huastecs, Notes on Middle American Archaeology and Ethnology, Carnegie Institution of Washington, Department of Archaeology, 114.223–227 (1953).



MAP A

Approximate geographical location of the six Mazatec villages under discussion

the longer ago historically the dialects separated, the list in Table 1 suggests the historical order of development of the Mazatec dialects. The comparisons fall too close together, however, for the implied exact order to be accurate in a statistical sense, since the calculated range of probable error of some items overlaps others. It would not be accurate, for example, to say that the separation of the Jalapa dialect from the San Miguel dialect certainly occurred earlier than the separation of the Jalapa dialect from the San Mateo dialect; the Jalapa dialect shares only ½% more cognates with the San Miguel dialect than with the San Mateo dialect, and the overlapping range of calculated probable error, 2.6% for each figure, makes it almost—not quite—equally probable that the reverse order of development was true.

In Table 1, however, there are four groups or clusterings of comparisons which are separated by gaps of 2½% or more. These groups were taken as the basis for study; they seemed significant in that, by and large, they were substantiated by independent comparative study[5] of these same dialects, which showed these groupings to be somewhat solid in a 'family tree' sense.[6] Also, although the borders of the groups

[5] A part of this tentative comparative study has appeared in, Sarah Gudschinsky, Proto-Mazateco, Memoria del Congreso Cientifico Mexicano 12.171–73 (1953). A more detailed study involving additional dialects is now under way.

[6] In an attempt to provide data with a smaller percentage of probable error, an additional list of sixty-four non-cultural words which were available in most of the dialects was added to the 200 word list, and the time depth was computed again on the basis of the enlarged list. This did not, as hoped, increase the spread of the comparisons. The significant fact, however, was that both with the original list and with the enlarged list, the comparisons fell into the identical groups, except for two minor variations within groups.

TABLE 1—Pairs of dialects listed according to the percentage of shared cognates*

No. of cognates	No. of words available for comparison	Percent of cognates	Pairs of dialects
		%	
138	185	74.5	Huautla–Jalapa
117	149	78.5	Huautla–Ixcatlán
138	172	80.0	Jalapa–Soyaltepec
146	181	80.5	Huautla–Soyaltepec
109	133	82.0	Jalapa–San Miguel
119	145	82.0	Jalapa–Ixcatlán
113	137	82.5	Jalapa–San Mateo
101	119	85.0	Ixcatlán–San Miguel
120	141	85.0	Ixcatlán–Soyaltepec
105	121	86.5	Ixcatlán–San Mateo
113	130	87.0	Soyaltepec–San Miguel
162	183	88.0	Soyaltepec–San Mateo
109	117	93.0	San Miguel–San Mateo
131	140	93.5	Huautla–San Mateo
129	137	94.0	Huautla–San Miguel

* The top of the chart implies the greater time depth.

show overlap of range of error calculated at higher confidence levels, the centers or means of the groups are clearly separate at $\frac{7}{10}$ confidence level, and the material that follows is in terms of the means of the groups of comparisons, rather than in terms of individual comparisons.[7]

[7] The development of a language is a continuum, and it is to be expected that dialect changes overlap each other, but that the center or crest of

1. The dip. At this point the usual formula of lexicostatistical dating was modified slightly so as to produce conclusions in terms of dips (i.e. degrees of lexical relationship), in order to avoid a misleading implication of time depth more certain than was justified by the data, in view of the overlap of the borders of the groups. Later on, where we found evidence of time-depth distortion because of reborrowing, the dips continued to prove useful in providing an objective basis for considering the distortion.

The term dip is derived from 'degree of relationship' as 'bit' is derived from 'binary digit'. The dip is computed by the formula $d = 14 \frac{\log C}{2 \log r}$ in which C is the percentage of basic vocabulary in common and r is a constant based on the estimated percentage of vocabulary retained after a standard period of time.[8] A separation of 2½ to 3 dips is necessary to avoid overlap of range of probable error (i.e. at 5/10 confidence level) if a list of 200 words is used.

In Table 2 the dialect comparisons are listed in terms of dips. The dips may be interpreted as expressing relationship, e.g., the San Mateo dialect is twice as close to the San Miguel dialect (3.3 dips) as to the Ixcatlán dialect (6.6 dips). Or they may be thought of as RELATIVE (not absolute) time. The San Mateo dialect has been diverging from the Ixcatlán dialect twice as long as it has been diverging from the San Miguel dialect.

Period I, as seen in Table 2, is a period in which the six dialects studied here are assumed to have been a single homogeneous speech community, in which we assume there were some dialectal variations, either phonological, or involving some part of the lexicon other than that included in our sampling, so that the variation does not show up in our data. Period I is inserted as an essential starting point in the discussion of the development of the lexical dialects, as we shall see later.

Period VI similarly refers to a comparable period of homogeneity, from a lexicostatistical view point, WITHIN ANY ONE of the defined speech communities, as it is now, at the other end of the relative time scale. It is listed as an essential point of reference for interpreting dips as relative time depth.

Period II, then, is 13.7 dips removed from Period VI; Period IIIA is 11.2 dips from Period VI;[9] the mean of Period IIIB is 9.5

---

the development be clear. For this reason the overlap of the borders of the groups is disregarded and they are related to each other in terms of their centers.

[8] The value r = .86 was used in the computations made in this study. The figure 14 in the formula was chosen in order that 14 dips would be equivalent to the first 1,000 years of time depth. This number of dips per thousand years was chosen for its empirical usefulness within this particular set of data. If by any chance it should prove to be empirically useful for other languages similarly closely related, that would be gratifying. On the other hand, further research may indicate that a larger or smaller unit would be more generally useful, in which case the formula can be readjusted to provide a dip of maximum usefulness.

The mathematical bases for the formulas used in lexico-statistical dating are discussed in Robert B. Lees, The Basis of Glottochronology, Lg. 29.113–127 (1953).

For discussions of lexico-statistical theory and its application to various problems, see: Morris Swadesh, Salish Internal Relationships, IJAL 16.157–67 (1950), and Diffusional Cumulation and Archaic Residue as Historical Explanations, SJA 7.1–21 (1951); and Joseph H. Greenberg and Morris Swadesh, Jicaque as a Hokan Language, IJAL 19.216–22 (1953).

[9] The division of Period III into sub-periods was done in order to partially simplify the maps used in Section 2 of this study. A second possible division of this period was not used because, in the study with the enlarged list (see footnote 6) this break within the group disappeared.

In Period IV there is also a break in the continuity of the group which might have been used as the basis of a subdivision. This break, however, did not mark anything useful in terms of simplifying the maps, and the subdivision could not be usefully exploited.

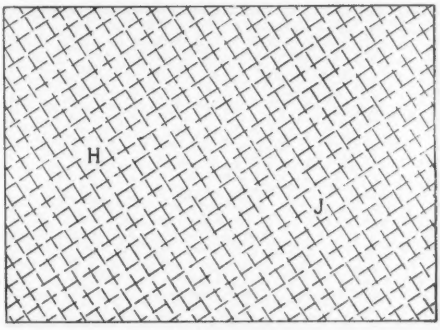
TABLE 2—Pairs of dialects at their respective separation points on the relative time scale in dips

Dips	Pairs of dialects	Period	Dips of periods
14	Homogeneous dialect area	I	14 or more
13.7	Huautla–Jalapa	II	13.7
11.2	Huautla–Ixcatlán	IIIA	11.2
10.2	Jalapa–Soyaltepec	IIIB	9.5
9.9	Huautla–Soyaltepec		
9.3	San Miguel–Jalapa		
9.2	Ixcatlán–Jalapa		
9.0	San Mateo–Jalapa		
7.6	San Miguel–Ixcatlán	IV	6.8
7.5	Soyaltepec–Ixcatlán		
6.6	San Mateo–Ixcatlán		
6.5	San Miguel–Soyaltepec		
5.9	San Mateo–Soyaltepec		
3.3	San Miguel–San Mateo	V	3.1
3.1	Huautla–San Mateo		
2.9	Huautla–San Miguel		
0.0	Homogeneity within each dialect	VI	0.0

dips from Period VI; the mean of Period IV is 6.8 dips from Period VI; and the mean of Period V is 3.1 dips from Period VI.

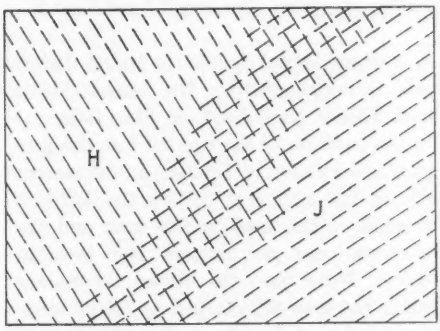
2. Lexical dialects. In **1** the periods of lexical development of the dialects were discussed with reference to relative time relationships in terms of dips. Now, on the basis of these same data, these periods can be mapped in terms of the lexical dialects which are clearly or partially differentiated.

Period I, already described as one in which the speech area is assumed to be homogeneous, is seen on Map B-1. The complete overlap of the down-slant lines, used on these maps to indicate the Huautla dialect, and the up-slant lines, used to indicate the Jalapa dialect, indicates the lack of lexical differentiation within the list of words which are included in our sample; the symbols H and J indicate the centers of assumed latent dialects which will show up in Period II as lexical dialect poles, since at that period there is differentiation between the Huautla and Jalapa dialects as seen in Table 2. It is the necessity for reference to this latent state which forced us to set up Period I.



MAP B-1, of Period I

A postulated homogeneous speech community, with latent dialect differentiation of the poles indicated by the complete crosshatching of their respective symbols: \ Huautla dialect, / Jalapa dialect.



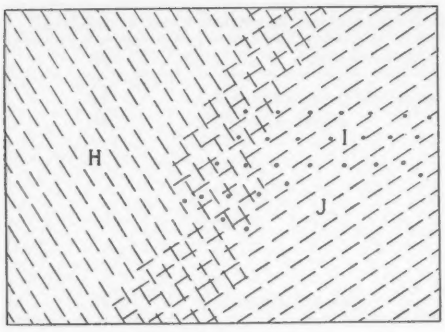
MAP B-2, of Period II

Here two distinct lexical poles have developed, symbolized by \ Huautla, and / Jalapa, with undetermined borders symbolized by the crosshatching of the symbols in the center.

In Period II, the data, as listed in Table 2, show that the lexical differentiation had progressed to the point where it can be detected by the lexico-statistical techniques, and hence Huautla and Jalapa are treated as distinct lexical poles with a transition area between them, whereas in Period I any lexical differentiation which was present was in peripheral vocabulary and not included in our sampling of core vocabulary.

Note on Map B-2 that the Huautla dialect continues to be symbolized by down-slant lines on the west side of the map, and the Jalapa dialect continues to be symbolized by up-slant lines on the east side of the map, but the Huautla symbols have been partly withdrawn from the Jalapa area and the Jalapa symbols have been partly withdrawn from the Huautla area, leaving the poles in contrast. The continued overlap of the two in the middle of the map indicates the postulated transition area where the borders of the dialect are still undetermined at this stage.

The next development, as seen at Period IIIA on the relative time scale of Table 2, is the differentiation of the Ixcatlán dialect from the Huautla dialect. Since the Ixcatlán dialect is not yet at this period separate from the Jalapa dialect, it is assumed to



MAP B-3a, of Period IIIA

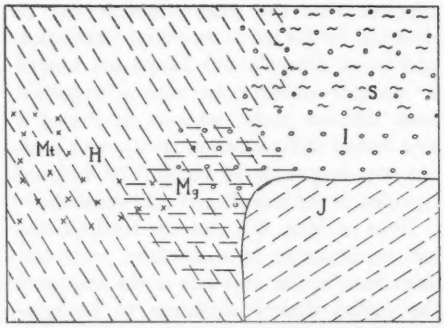
The emerging of a lexical pole within the Jalapa area is marked by the symbol O Ixcatlán, and identified by the lexical differentiation of the Ixcatlán dialect from the Huautla dialect later than the differentiation of the Jalapa dialect from the Huautla dialect. (See Map B-2 and Table 2).

be a latent lexical dialect, or lexical dialect pole, developing within the Jalapa dialect area. This is symbolized, on Map B-3a, (a) by the addition of the symbol I at the geographical location of Ixcatlán on the eastern side of the map, and (b) by the addition of small circles, used in this map and the following ones to indicate the Ixcatlán dialect. Note that the circles lie mainly within the area still marked as Jalapa territory by the up-slant lines, since the dialects of Jalapa and Ixcatlán cannot yet be distinguished at this period by lexico-statistical techniques. There is some overlap of the circles with the down-slant lines to show that, although the Huautla and Ixcatlán dialects are lexically distinct at this period, the location of the border between them is as yet uncertain. For the same reason, the overlap of the up-slant and down-slant lines remains, marking a transition area between the centers of the Huautla and Jalapa dialects.

Period IIIB includes a clustering of dialect separations. The separation of the Soyaltepec dialect from both the Huautla and Jalapa dialects, as seen in the lexicostatistical data in Table 2, indicates the emergence of this dialect as a new lexical pole within the Ixcatlán dialect from which it is not yet separate. Similarly, the fact that the San Miguel and San Mateo dialects are, at this period, separate from the Jalapa dialect, indicates their emergence as new dialect poles within the Huautla dialect, from which they are not yet separate.

This period also marks the completion of the definition of the borders of the Jalapa dialect, as we can tell from the fact that in Table 2 the Jalapa dialect is shown to have separated from each of the other dialects in turn, so that by the end of this period it is separate from all of the other five dialects considered in this study.

This definition of the borders of the Jalapa dialect in turn leaves the Ixcatlán area separate from both the Jalapa and Huautla areas. In Map B-3b the developments of Period IIIB are symbolized: (a) by the addition of the symbols S, Mg, and Mt at the geographical locations of Soyaltepec, San Miguel, and San Mateo; and (b) by the addition of tildes to indicate the Soyaltepec dialect, crosses to indicate the San Mateo dialect, and dashes to indicate the San Miguel dialect. Note that the tildes lie mainly within the area still marked as



MAP B-3b, of Period IIIB

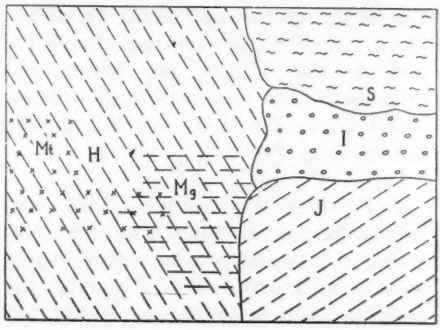
The development of lexical poles within the Huautla and Ixcatlán dialect areas, and definition of the borders of the Jalapa dialect, O Ixcatlán, ~ Soyaltepec, × San Mateo, and — San Miguel.

Ixcatlán territory by the circles, since the dialects of Soyaltepec and Ixcatlán cannot be distinguished at this period, although both are lexically separate from Huautla. The upslant lines marking the Jalapa dialect, however, have been completely withdrawn from this area, since the Jalapa dialect is now separate from both Ixcatlán and Soyaltepec. Note also that the crosses and dashes lie mainly within the area still marked as Huautla by the down-slant lines, since the dialects of San Miguel, San Mateo, and Huautla cannot be distinguished from each other at this period, although all three are lexically distinct from Ixcatlán and Jalapa. There is some overlap of the down-slant lines, circles, crosses, and tildes to show that the location of the borders of these dialects is still uncertain. The up-slant lines of Jalapa, however, no longer overlap with any of the others, and the solid line surrounding this area indicates that, within the limits of this study, the Jalapa dialect is not further subdivided or differentiated.

Period IV, as seen in Table 2, marks the completion of the definition of the borders of the Soyaltepec and Ixcatlán dialects, in that each of these has now appeared in the data as separate from each of the other dialects in turn. In Map B-4 the developments of Period IV are symbolized by the addition of solid lines around the Soyaltepec and Ixcatlán areas to indicate that, within the limits of this study, these two dialects are not further subdivided or differentiated. Note, however, that the crosses, dashes, and down-slant lines continue to overlap, indicating that the joint borders of Huautla, San Miguel, and San Mateo dialects are still undetermined.

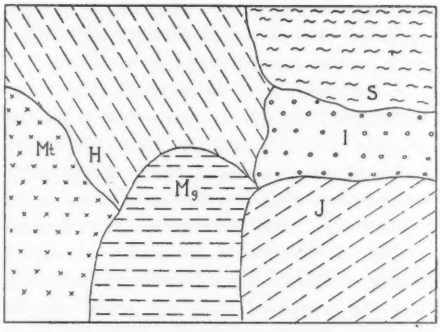
In Period V, as seen in Table 2, the remaining dialects are seen as separated from each other, indicating the completion of the differentiation of the joint borders of the San Mateo, San Miguel, and Huautla dialects, so that the six dialects have achieved the total separation indicated in Period VI. On Map B-5 this final differentiation is symbolized by the addition of solid lines around Huautla, San Miguel, and San Mateo. This map now represents the six Mazatec dialects which we are studying, as they are today, and the set of maps as a whole indicates, by the successive appearance of each solid line, the relative time of the complete differentiation of each dialect from all of the others.

The Jalapa dialect was completely differentiated during Period IIIB, approximately 9.5 dips removed from Period VI. The Ixcatlán and Soyaltepec dialects were



MAP B-4, of Period IV

At this stage, the complete definition of the borders of the Soyaltepec and Ixcatlán dialects has occurred, but with continued transition area within the Huautla dialect area.



MAP B-5, of Period V

The definition of all six dialects under consideration is complete by this period.

completely differentiated during Period IV, about 6.8 dips removed from Period VI; and the San Miguel, San Mateo, and Huautla dialects were completely differentiated during Period V, only 3.1 dips removed from Period VI.

3. Phonological dialects. Having considered the lexicostatistical data as an indication of historical periods in the development of the lexical dialects of Mazatec, the question arises as to whether or not this scheme of development is confirmed by the periods of development deduced from sound correspondences of these dialects as studied by the traditional methods of comparative linguistics.

To answer this question, thirteen sets of correspondences were studied from the point of view of sound changes with reference to the reconstructed proto-forms.

I assumed that, in general, the more widespread changes in these closely related dialects occurred earlier than the more localized sound changes which are peculiar to one area. This assumption allowed us to tentatively map successive phonological dialect developments in terms of sound correspondences. These developments could then be compared with the lexical dialect developments as mapped in **2**, on the basis of lexico-statistical dating.

Of the thirteen sound changes studied, the most widespread are a group of four changes in which initial *C is lost from clusters *CCC. This group of changes affects the dialects of Soyaltepec, Ixcatlán, Jalapa, and San Miguel.

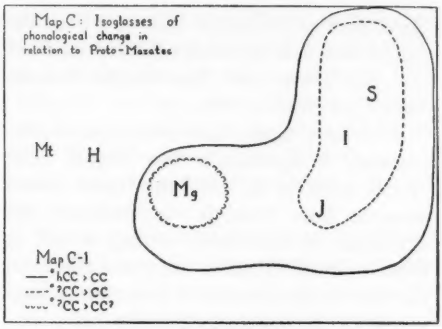
The loss of initial *h and *ʔ is shown on Map C-1. On this map, the solid line encloses the area in which *hCC > CC as in: hnku *one* > nku in Mg, I, S, and J. The line of dashes encloses the area in which *ʔCC > CC as in: *ʔnči *wet* > nči in S, I, and J. The small curved lines enclose the area in which *ʔCC > CCʔ as in: *ʔnči > nčʔi in Mg. Note here that loss of initial *ʔ occurs in the lowland dialects and San Miguel; but may be considered incomplete in the latter in that it yields metathesis rather than complete loss.

These phonological data, like the lexical data of **2**, and specifically Map B-2, indicate an early dialect separation between Huautla and the group of non-Huautla dialects, with some differentiation within a transition area between the two. Note, however, that San Miguel, which showed up as being lexically close to the Huautla dialect in **2** (see Maps B-3b and B-4), is here definitely closer to the non-Huautla dialects—a fact which will prove very important in the remainder of this paper.

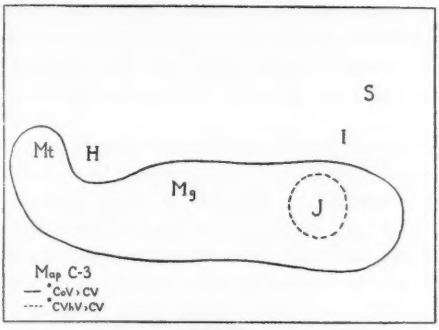
The loss of initial *n is seen on Map C-2. On that map the solid line encloses the area in which *n is lost from clusters of *nCh, as in: *ncha *hair* > cha in Mt, Mg, S, and I. The line of dashes encloses the area in which *n is lost from *nCʔ, as in: *nčʔą *cold* > čʔą in S. The dotted line and the line of curves around Jalapa indicate that, in that dialect, the loss of *n from the clusters of *nCh and *nCʔ is incomplete, so that there is morphological alternation of ncha ∞ cha *hair* and nčʔą ∞ čʔą *cold*. The line of crosses encloses the area in which *čh > ç̌h as a result of the loss of *n from the clusters of *nčh as in: *čhų *woman* > ç̌hų in S and I.

Note in this map the agreement with Map B-2 in the development of dialect poles at Jalapa and Huautla. Note also that the development of differentiation of phonological dialects within the dialect transition area agrees to some extent with the order of lexical development suggested by Maps B-3a,b and B-4. For example, on Map C-1, an over-lapping area includes Ixcatlán, Soyaltepec, and Jalapa as on Map B-3a, and on Map C-2, the further differentiation and separation within this area corresponds to the lexical developments shown in Map B-3b. This order of events in the phonological dialects is postulated as the result of the application of our assumption that the more widespread changes were earlier changes.

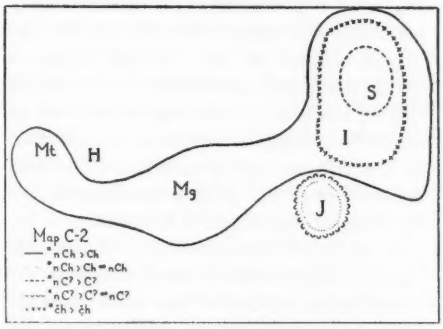
Map C-3 represents two other sound



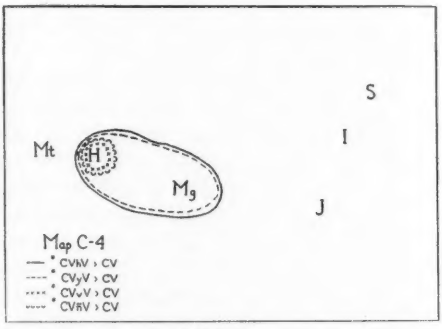
MAP C-1



MAP C-3



MAP C-2



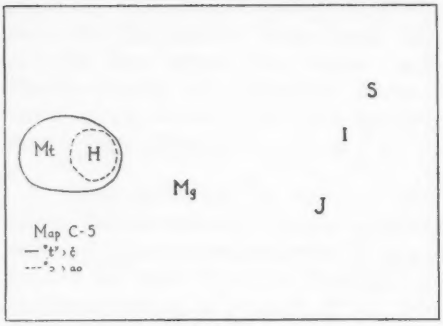
MAP C-4

changes. One, the loss of *o following C_2 (*C_2 being consonant other than *k) is shared by Jalapa, San Miguel and San Mateo. The solid line on the map encloses this area in which *C_2oV > CV as in: *choa *happy* > cha in Mt, Mg, and J. The other, the loss of *V, is peculiar to Jalapa. The dotted line encloses the area in which *CVhV > ChV as in: *ntoho *long* > ndho in J.

Note on this map the further differentiation of Jalapa, paralleling its lexical differentiation as seen in Map B-3b. Also note, again, the phonological development of San Miguel dialect as part of the Jalapa dialect area, in direct contrast with its lexical development as part of the Huautla dialect area as seen in Maps B-3b and B-4. This continues to point up the chief problem in correlating lexical and phonological developments which we have hinted at a number of times, as for example in describing the loss of *h above.

Four sound changes form a group of changes involving loss of medial *C from *CVCV. Two of these four losses are peculiar to Huautla, as seen in Map C-4. The line of crosses on this map encloses the area in which medial *v is lost, as in: *sava *shame* > soa in H. The small curved lines enclose the area in which there is loss of medial *ñ, as in: *tiña *near* > čią in H. The solid line encloses the area in which there is loss of medial *h as in: *ntoho *long* > nto in H and Mg. The line of dashes encloses the area in which there is loss of medial *y as in: *ntiya *road* > ntia in H and Mg.

Since these four changes seem to be aspects of a single change (that is, loss of medial *C), there is reason to suppose that the innovation occurred in Huautla, where



MAP C-5

all four changes occur, and spread to San Miguel recently through borrowing.

A similar hypothesis may[10] account for the fact that one sound change is shared by Huautla and San Mateo. In Map C-5, the solid line encloses the area in which *tʸ > č as in: *tʸha *arm* > čha in H and Mt. It is assumed that this change was initiated in Huautla and spread to San Mateo. One other change is indicated on Map C-5, on which the dotted line encloses the area in which *ɔ > ao as in: *hɔ*two* > hao in Huautla. Note that this change is peculiar to Huautla. It is presumably a development subsequent to the development of other vowel clusters following the loss of medial *C.

4. Discrepancy between lexical and phonological dialects. The history of the phonological dialects may be summarized as (1) separation of the Huautla dialect from the other dialects, (2) differentiation of the non-Huautla dialects, and (3) further development of the Huautla dialect and borrowing from Huautla by San Miguel and possibly by San Mateo. By contrast, the lexical data is summarized as: (1) early differentiation of the Huautla and Jalapa dialects, (3) later differentiation of the Soyaltepec and Ixcatlán dialects within the Jalapa area, and (3) finally, differentiation of the San Mateo and San Miguel dialects within the Huautla area.

Note here again that the point of discrepancy is specifically San Miguel. The crucial question is, did San Miguel dialect separate from Huautla very recently (as suggested by the lexical study) or did it separate from Huautla early and then redevelop in the direction of Huautla (as suggested by the phonological study)?

I postulate that the dialect of San Miguel separated from the dialect of Huautla early, and that the lexical closeness of the San Miguel dialect to the Huautla dialect is due to recent lexical reborrowing. I postulate this on the basis of documented cases of recent reborrowing, the geographical proximity of the villages, the economic and political dependence of San Miguel on Huautla, and the cultural dominance of Huautla.

A study of the phonemics of the San Miguel dialect reveals a small sub-system of words which do not fit the main phonemic system of the dialect, or which do not follow the regular historical development of other words of the dialect.

In every such case, the forms are either identical with Huautla forms which are phonemically normal parts of the Huautla system, such that one can deduce that they are borrowed from the Huautla dialect; or they are clearly derived historically from Huautla forms rather than from the proto-forms.

Among these borrowed words are a group of religious terms such as nʔai *father* and nai *devil* which have vowel cluster ai which does not occur elsewhere in the dialect. The word čʔa *parrot* (also a religious term, since the feathers of this particular type of parrot are used in witchcraft), exhibits historical development from the Huautla form čʔoa by loss of o, rather than the expected correspondence with the proto-form: *čoʔani > čoʔani.

The lexical borrowing of San Miguel from Huautla is facilitated by the geographic

[10] The San Mateo dialect is phonologically the most conservative of the six dialects studied. Within the data considered here, it is not certain whether San Mateo, like San Miguel, has reborrowed recently from Huautla, or whether it is part of a different dialect area including other villages located north and west of Huautla.

proximity of the two villages as described in **0**, and also by wide, easily traveled trails; San Miguel lies on the main route from Huautla to two other Mazatec villages. Huautla in turn lies on the main road from San Miguel to the nearest Spanish market town. Merchants and mule drivers make this trip regularly.

San Miguel is economically and politically dependent on Huautla. Although it has, technically, the status of an independent township, the more important matters, such as the investigation of a murder, are frequently handled through the Huautla town government. In addition to this, Huautla is a large market town and dominates the local trade. Coffee beans, the main cash crop of the area in general and of San Miguel in particular, are sold almost exclusively through Huautla outlets.

Culturally also, Huautla occupies a place of dominance. I have frequently observed the men of San Miguel making every effort to speak the Huautla dialect, and have seen them pretend to be natives of Huautla until someone pointed out dialect characteristics of San Miguel, at which point they shamefacedly admitted they were from San Miguel. The present observable tendency in the village is in the direction of deliberate adoption and cultivation of the Huautla dialect, especially in town business and in trade.

5. Conclusion. This evidence leads to the following analysis of the Mazatec dialects: (1) an early period in which there were lexical and phonological differentiation of Huautla from the other areas; (2) next, a period in which there was lexical and phonological differentiation within the non-Huautla dialects, and (3) a later period during which there was a great deal of lexical borrowing by San Miguel from Huautla.

Now, we are prepared to answer the question raised in **4**, by saying that the lexico-statistical data by itself gives a false picture of the relationship of the Huautla and San Miguel dialects, so that the historical perspective is skewed in that the Huautla and San Miguel dialects appear to have separated more recently than is the case.

Or, viewing this same data from the point of view of Jalapa rather than Huautla, the skewing could be restated as follows: the historical perspective implied by the lexico-statistical data is skewed so that the Jalapa and San Miguel dialects appear to have separated much earlier than was actually the case. This skewing, however, can be partly corrected in terms of the phonological data, which gives the truer picture of the historical sequence in which the dialects of the area were differentiated. Nevertheless, in spite of the fact that the distortion can be corrected in terms of relative sequence, or relative time, any hope of determining absolute time—by some improvement in the precision of lexico-statistical technique—seems to have been permanently lost, since the order of differentiation can be deduced from the comparative material, but the precise number of cognates lost and reborrowed cannot be so deduced.

What, then, are the implications of this for the lexico-statistical techniques? We conclude that lexico-statistical materials lead to very useful implications as to the historical sequence in which dialect differences developed, but that reborrowing from other dialects following a change in the lines of communication or a shift of cultural and economic dependence, can distort the evidence so that it no longer accurately reflects the true historical development.

When, therefore, no comparative data are available as a check on the reliability of the time depths suggested by the lexico-statistical dating, we find the results less subject to uncritical acceptance by the unwary if we deal in relative degrees of lexical relationship, expressed in dips, rather than in terms of absolute time. In addition, the current degrees of differential lexical relationship can be objectively described without a time bias, by means of dips

MOENCOPI VARIATIONS FROM WHORF'S SECOND MESA HOPI

CLYDE KLUCKHOHN
HARVARD UNIVERSITY
AND
KENNETH MACLEISH
NEW YORK CITY

1. Whorf's letter to Kluckhohn
2. Whorf's 1937 Hopi Phonemes
3. Moencopi variations

1. WHORF'S LETTER TO KLUCKHOHN (DATED JUNE 9, 1937)

Dear Dr. Kluckhohn: I am sending you today or tomorrow the Hopi vocabulary which I have prepared for you. I got interested in the job and it is somewhat larger than I at first intended, but should be the more useful for that reason. I am rather interested to see what will happen with an ethnologist going to the field already equipped with a substantially correct vocabulary and outline of the language of size sufficient for all conversational purposes. The words are on cards, and these are divided into the sets: nouns, verbs, adjectives, adverbs, postpositions, pronouns, temporals, numerals, interjections, verb suffixes. The use of cards of course also enables you to classify the nouns or verbs ethnographically or in any desired way. Each set is arranged in the order of the English alphabet, with modified characters like 'p placed after those on which they are based. Exceptions to this are: ʔ is placed as the first letter of the alphabet—the order of the vowels is a, ɛ, e, i, o, ö—e is used throughout for ə, the usual phonetic symbol, which is employed in the list of phonemes I sent you. I recommend that e be used for this phoneme, and made standard. I have marked k before a and ɛ as k̡—this is not necessary and the mark should be dropped in print—it is placed there merely to remind you that this is not the English k in this position, but sounds like ky.

Wherever possible, the ordinary isolation form of each noun or verb is followed by the base to which most suffixes are attached, or in the case of nouns, the base in which the noun is used as first member of a compound. Thus the verbs ʔe'ʻta, wa'ri, and nö·'sa receive the future suffix -ni respectively as ʔe'ʻtàni, wa'rïkni, and nö'sni; this is indicated by citing their bases as ʔe'ʻtà-, wa'rïk-, nö's-. The base is indicated by a hyphen following. As far as possible the usual plural of each noun is given, since there are several types of plural formation. Many words have two forms, a longer pausal form used at the end of a sentence and a short form within the sentence. This is indicated by putting in parenthesis the part of the word which is dropped in the short form, e.g. ma'mant(e) 'girls'. Thus the vocabulary as prepared enables you to short-cut some of the work that requires a long time in digging out these basic grammatical facts from informants.

The list of suffixes is also intended to give you a good approach to the fundamentals of the grammar and enable you to see through things that otherwise might be very puzzling. Use the clear spaces on the cards for additional notes, and give the set back to me in the fall. It may not be necessary to have it typed.

I am interested in the fact that you now have linguistic equipment for getting facts in the native tongue such as is rarely found among American ethnographers. Ability to speak Hopi moderately well, using the simpler types of sentences, should not be long in coming. If you have time, I should strongly advise re-indexing the vocabulary according to the English meanings, making up an English-Hopi manual. This will give

you a great command of expression. I hope that the many interesting forms and meanings of these Hopi words will intrigue your interest and stimulate your desire to make through linguistics a better approach toward understanding the native culture.

Before you leave Harvard will you please have Peabody Museum send me 10 copies of my paper published by them "The Phonetic Value of Certain Characters in Maya Writing". My set of reprints is exhausted and people keep writing to me, rather than to the Museum, for copies. Have them bill me if there is any charge.

Wishing you a very successful field trip, Sincerely yours, B. L. Whorf.

2. Whorf's 1937 Hopi Phonemes [second mesa]

2.1. Primary consonant phonemes

Stop series	Allophones—positional variant phones and remarks
p	p
t	t alveolar (i.e. like English t)
c	syllable-initial = tš^{s} (between tš and ts); syllable-final = ts
k	k͡a, k͡ɛ = k^{y}a, k^{y}ɛ ki = Eng. ki (keen) kə, ko = Eng. kə. ko (curl) −k (syllable-final) = Eng. −k (back)
ķ	Only as ķa in Spanish loan words (ķawa'yo, ķanɛ'lo, ķa'stila etc.) Like Eng. ka- (calm)
q	Only inqa-, qɛ, qö-, -q (syllable-final) velar q, like Arabic or Nootka q. Between vowels like -qx-
k^{w}	Eng. k^{w} (queen)
Nasal series	
m	m
n	n
ŋ̍	ŋ^{y}
ŋ	ŋ (in Eng. hang, but more velar)
ŋ^{w}	ŋ^{w} only in ŋ^{w}- (syllable-initial) and ŋ^{w}ʔ-
Continuant series	
v	v− = β− or v− −v = −ɸ or f −vv = −ɸβ or −fv−
s	s or s^{š} (slightly toward š) The position between s or š depends on the dialect.
l	l (as in English)
r	r untrilled, retroflex, with slight friction suggestive of ž. Never write this phoneme as rš, zr, ž, z, etc.
Semivowel series	
w	w as in Eng. w−, −w
y	y as in Eng. y−, −y
h	h as in Eng. h−, −h
ʔ	ʔ glottal stop. Never ɣ in writing. Every Hopi syllable and word begins with consonant phoneme. Hence never . . . a word beginning with a vowel—not anca, o'va, but ʔa'nca, ʔo'va.

The semivowel series gives more trouble to the beginner than other consonants. All apparent falling diphthongs e.g. aɪ̯, oi̯, aʊ̯ or aǫ̯, iʊ̯, or iǫ̯ etc. are vowel plus consonant; ay, oy, aw, iw etc., and occur in all possible combinations, e.g. əy, əw (sometimes taken for an u), öy etc.; iy must be distinguished from i. All apparent rising diphthongs, e.g., i̯a, ʊ̯a or ǫ̯a are similarly -ya, -wa etc. *They sing* is ta·'wlálwa, not ta·'ylàloa. Word-final -ʔv (phonemically) is phonetically -ʔvh—v = any vowel; c = any consonant.

ʔ is tinged with the quality of the neighboring vowels, which is especially apparent with ə and ö—do not write ɣa for ʔə or yö for ʔö—*salt* is ʔö̍'ŋa, not yü'ŋa or yz'ŋa. *Buys it* is tə̍'ʔi, not tə̍ɣi. There is no ɣ (gamma, voiced guttural spirant) in Hopi, though written by some anthropologists.

Clusters of three consonants cannot occur within a word unless one (or more) of the three is of the semivowel series.

The combination -ʔyc is common and rather hard to hear, nö·'màʔyta *has a wife* is apt to be heard as nö·'màʔta or nö·'màyta.

2.2. Secondary consonant series.

Preaspirated stop series: 'p, 't, 'c, 'k, 'q, 'k^{w}. The corresponding stops preceded by a soft aspiration, a little softer than h, and apt to be tinged with the quality of the

preceding vowel, e.g. a velar quality after ə. *Woman* is wə''ti, not wəɣti, wəxti, or wəqti. This series is not found in some dialects, e.g. Shipaulovi, but is represented by lengthening of the preceding vowel, e.g. po·'ta and po''ta are both po·ta in Shipaulovi dialect.

Desonantized (unvoiced) series: M, N, N̲, N̲̲, L, R, W, Y. Like the corresponding sonants m, n, etc., but unvoiced. Occur only syllable-final, but the sonants occur in the same position and must be distinguished: ʔa'nta, pą'Nti; *if* is kǫR, not kǫs or kǫs.

2.3. Vowel allophones

Vowel Phonemes	Bright	Dull
a	a (calm)	ɑ (but)
ɛ	æ ä (bad)	ɛ (met)
i	i (keen)	ɪ (kit)
ə (e in my orthography)	ə (burn)	ĭ after labial (further back than ə) ɪ further front than ə— similar to Polish y or Welsh u
o	o, oų (note)	ʊ (put)
ŏ	ŏ fleuve (French) but before ŭ (German Mutter, Grunde) q almost ɔ (log)	

English speaking anthropologists have been prone to confuse Hopi ə and ŏ. Here are some test words often wrongly recorded: ʔə''ta *opens it*, hŏ'ta *shuts it*, hə'wi *trap*, hŏwi *dove*, hə'mi *rubbed-off corn kernels*, hŏ'mi *hair*, kə'yi *water*, gŏ''hi *fire*, lŏ'yə *two*, lŏ'wa *vulva*, mə·'yâw *moon*, pə·'ka *skin*, nŏ·'sa *eats*, tə·'kʷi *mountain*, tą'ckʷa *ground*, cə'ʔa *rattlesnake*, cŏ'viw *antelope*, wə''ti *woman*, wə'ŋʷa *clan-affiliation*, wə·'wəcìm *name of adult-initiation society and ceremony*.

Rules for the vowel allophones:

a (1) when short (ą) and not before or after g, is dull; (2) in all other situations, is bright.

ɛ (1) when long (ɛ·) is bright; (2) otherwise, is dull.

i (1) in a closed syllable, and when short after w, kʷ, ŋʷ is dull (more or less); (2) otherwise, is bright. In both cases, before q is followed by a glide ə; e.g. niq = niəq.

ə (1) when short (ą) is dull, and after a labial has the far-back, guttural ĭ sound, otherwise more or less the i-sound; (2) otherwise, is bright.

o (1) when short (ǫ) is dull; (2) otherwise, is bright.

ŏ (1) when short (ö) is dull, also in closed syllable usually dull; (2) otherwise is bright, more or less. See table above regarding before q.

ą (short, stressed a) after y, k, ŋ is almost ä. Thus, heard yąæ = yą, heard yä· = yɛ·. heard yɛ = yę.

2.4. Stress and length phonemes.

2.4.1. Stress

Stressed syllables:
Primary stress a′
Primary stress falling â
Secondary stress à
Unstressed syllables: unmarked

2.4.2. Vowel-length

Vowel length distinctions apply only in stressed syllables: short ą, medium (half-long) a, long a·.

Vowels in unstressed syllables are about medium in length or slightly shorter. They do not count as short in phonetic rules.

Some words from a text illustrating stress and vowel length: ʔo·'viy ʔə'yə̀wsiy kʷa·'vokmą̀t tama'latą̀ni tənö'svìsokʔìwta pą'ŋqàwə po'ʔoLtinǫ́qat ʔi·'kʷìwkaŋ po''komą̀t ʔi·'kʷiwnǫ́ma ʔöma''totą'tsiwa'ʔatyori'kt yǫ'ri ʔamǫ'mi ŋəʔa'yą̀ni wə·'wa'ya ta·ýyə̀ŋʷa wikcây ʔɛ'cvɛtɔ̀nayanî

[See now, B. L. Whorf, The Hopi Language, Toreva Dialect, VFPA 6.158–83 (1946)].

B. L. Whorf
March 8, 1937

3. VARIATIONS OBSERVED AT MOENCOPI (1937 and 1938) FROM THE HOPI LINGUISTIC MATERIAL FURNISHED BY B. L. WHORF.

3.0. All of the points here set forth have been checked [by Clyde Kluckhohn and Kenneth MacLeish] with at least four informants, representing individuals from 24–65 plus years of age. They consist, in the main, of small phonetic variations in specific words and of semantic additions and qualifications. There are, however, four generalizations which may be made [by Clyde Kluckhohn]:

3.1. When it precedes the consonants t, q, h, v- is invariably *β*. (Probably true when preceding any consonant? My examples cover only these three cases.) Also as word-final = *β*.

3.2. Between k- and -p Moencopi speakers invariably insert -in-. Examples: pese'kinpi, not pese'kpi; ʔɛyo'kinpi, not ʔɛyo'kpi. I was told that this was characteristic of the speech of all Third Mesa villages. Similarly, perhaps: kʷɛci'n'pi, not kʷɛci''pi.

3.3. Frankly, I am unconvinced that Moencopi speakers at any rate distinguish PHONEMICALLY between short and medium length. The only example you [Whorf] gave me [Kluckhohn] was yo'wi *rain* and yọ'wi *corn-silk*. The former word, I am quite certain, is never used in Moencopi—yo'wki *rain* is used. I was unable to discover a single case in which the difference between short and medium (half-long) was distinctive.

3.4. The alternative plural form of nouns which gives a -t ending in place of a smooth vowel seems, in many cases at least, to convey a distributive sense. For example, ʔa·'ʔàca was said to mean *lies told by a single individual*, whereas ʔa·'ʔàcat was said to mean *lies told by more than one person*. On the other hand, this could hardly be the case with the common Moencopi plural of *horns* ʔa·'ʔalàt. I am, in general, much less confident of this uniformity than of the three previously set forth.

3.5. Henceforth I shall simply list words and state variations heard or semantic additions or qualifications:

ʔa'ʔne? (rather than -ɛʔ). Common meaning in M. *industrious*.

ʔana'haʔà Additional meaning *thorny*.

ʔani'Wpa the basic meaning, according to my informants, is *something that is raised or cultivated*. Can be—e.g.—*lettuce*. The word ʔani'Wti means *working at something and get it right*; also commonly used to refer to the years in which seasonal crops (pinyon, for instance) come in.

a''pa refers primarily to the sheepskin or pallet type of bedding, but is also sometimes used with reference to the bedding placed on a bedstead.

ʔa'cvèwa (rather than ʔɛ-). Reduplicated plural also based on -a-.

ʔeʔe'cpi The plural I could only hear as ʔe'ʔècpi.

ʔe'ya Plural in M. ʔe'eʔyat.

ʔe·'yi The plural was said to mean also *stolen goods*.

ʔị'Npi This is the common word for *shovel* in M. Also means *spade* and *plate*.

ʔo''va Also means *grape*. [In addition to Whorf's gloss]

ʔǫ̈'sö I tried fully 20 persons, men and women, on this, and all were doubtful as to the plant referred to.

canʷẹ' sisiva This is the singular in M. The unreduplicated form is apparently not used. The common plural is ca·caŋʷẹ' sisivàt.

caqa'vta Of anything big and somewhat spherical. Never used of a plate, for instance.

ca·'yo commonly occurs as an alternative form.

cị'ro Said to refer only to small birds, "especially winter birds that come while the snow is on the ground."

cẹ'ke Plural ce'cket

cọ'mo Plural co'cʔ mot (plurals ending in -t were usually given to me as a form preferred to smooth vowel ending; I have not

troubled to list all cases for the tendency is almost exceptionless).

ho'he Definitely in M. not *juniper* or *cedar* but *piñon*.

ho·'he Plural ho·'ho?at.

ho'ŋa My informants were unfamiliar with this non-collective form. They said it meant to them only *something that is standing up big and high*.

-hòya piya'Rhòya said to mean also *little goat* and to be used as a kind of nickname for little boys and little girls.

ka·'ro Plural -Rot.

ki·'si Plural ki·'kist.

kọ'hɛ In M. ka'pɛ or (?) ka'phe. Plural ka·'kàphe *sacks of coffee*.

kʷa·'he Used in M. but more common form is kʷa·'te.

kʷɛ'wa means *any kind of belt*.

kʷe'he Not used in M. Word used there is ko'ho.

kʷici'ŋʷe (not -ce-). The final vowel here, as in similar positions, is almost unvoiced, and one has to strain to hear the labialization.

kʷį'ŋavì My informants stated that this was a kind of *willow*, the wood of which is used for bows, whereas kʷi'ŋvi (they said) was a *scrub oak*.

lę'he This word was declared unknown, although lepọ'si was recognized.

lɛ·'na Any kind of musical instrument, also phonograph or radio. *Violin* is ta'he. lõha've (rather than -vi).

ma'nsi noy *flower* generally; a type of Indian paint brush, specifically *argenacia linareafolia*.

ma'ntɛwà (rather than -t-).

ma·'sàw (rather than ma-). Two informants independently volunteered observations of this sort: "In the other Hopi towns, they talk as if they were out of breath. Here everything is long."

masi'vi My informants agreed that this was a *grey fly* "which hangs around lamps at night." It is possible that I was being put off. On the other hand, the other words for *ghost* were admitted freely enough.

me?á'pi Also *nail* and *any kind of fastener*.

meca'pe Refers mainly to *canvas*. ḳanɛ'l —— means the heavy black woolen shawls worn by women.

me·'na Plural me·'menaNt.

mẹ're It was said that this form is not used independently in M. It occurs in various compounds such as name're *ridge*.

me·'yi Plural me·'meyì·m.'

mo·'ho (rather than -he); mo·'vi is *yucca root*.

mạ'sa *wing* only; *feather* is homạ'sa.

mọ'ri Plural mori'vosì.

mǫ̈'?wi- Plural mǫ̈'mö? wit.

na·'qö'yta Informants objected to this as a noun meaning *war*, saying it meant *killing each other*, while *war* (in the past) was na·ǫ́ö'ywe.

na·'sàvɛ (rather than -vi).

na·'soMpi any short rope or string.

na·'ve (often -vi).

no·'va no·'nòva is the *eating together of a group of people*.

ŋõ'la Not *circle* which is po'ŋòq'pe.

pa·'ki Not *sob* etc. but *boat—water-house*.

pa'ŋʷe plural pa·'vaŊwet.

pa·'qàla plural-qala.

pa·'sa *any cultivated place*.

patɛ'seŋʷa (rather than -e's-).

patɛ'vha (rather than -e'v-). Perhaps in this and the preceding example we are dealing with a uniformity which might have been generalized.

pa·'va Some say ?i'pavà rather than ?i'v-.

pica'ŋʷa (rather than -e). *Face* in the sense of distinctive features—not in the sense of a bodily part.

pika'?yŋʷa plural pi'vik-.

pi·'ki (commonly).

piqa'vikì plural (cf. above) pi'viq-.

poọ'si common word for *seed* is pọ'shemi.

po'ŋa plural po'vo-.

pö·'ca This word was not recognized by any informant. The word for *skunk* was said to be go'liciyö' (from Navaho?); the word for *caterpillar* pi'?ake

pọ'yo (First Mesa—po'yo? ???); plural (M) pọ'yot.

pota′sivè *basket* of roughly olla shape—not flat trays etc.
po·′ta Not a *plaque*; *shallow basket.*
powa′qa plural po′vow-.
pö′sa plural pö·′vöyam.
qa·′la *wood rat*, not *field rat.*
re″pi also: *medicine man's crystal.* [In addition to Whorf's gloss]
sa·′qa plural sa·′sàqa.
sipa′la *the downy peach.*
söhö′vi My informants did not agree that this was used for *tree* in general.
soso′Mpi plural so·′soMpi.
ta′ŋàqa My informants said this was not used by itself, but only in compounds.
tasa′ve The Navahos object to being called this; they prefer being called qalá ʔlta·′qa.
ta·′we My informants professed complete unfamiliarity with this term. *Heel* is commonly qe′ktinsi.
te·′kʷi Not *mountain* but *butte* or *mesa.*
te′ŋʷni (rather than -ŋni).
tepe′vi Word unknown in *M.*
te′vqa is a *deep, straight-sided canyon,* while si′ka applies to the *shallower canyons* with more broken rims and talus slopes.
te·′vè ni *paper, writing.* Only the plural means *letter.*
te′wiki *helper*—more permanent than ʔạya′ who is just for helping finish up a specific task etc.
tị′he almost exclusively restricted to the kachina doll.
tokọ′ci idiomatically used to refer to a quick-tempered person.
to′ni *thread, yarn, string.*
tö·′sö A common alternative form in M is p-.
to·′ci (First Mesa k-).
wa·′la mainly *arroyo, wash.*
yama′kinpi also: *ford* (crossing).

ADJECTIVES

ʔa′tkahàq I was told that this meant not so much *deep* as *down below* when pointing.
ʔɛ·′ka′wi Said to refer to animals or men "that move slow, never did go fast."
ʔici′ve *irritable, quick-tempered* rather than *angry.*
hi′sakìs Primarily an interrogative.
koko′ma Also a noun referring to a certain plant with flowers of this color.
mo′ki This simple form unknown in M.
nɛvẹ′ lvecqà *flat,* not *square.*
nikʷạ′ŋʔew More nearly, according to my informants, *advantageous.* They said it meant *making better use of time* etc. e.g. "getting a chance to go to Flagstaff with someone else." *Perfect,* they said, was ne′kŋʷa.
pɛto′ si *reddish-brown.*
pe′cqa Also used as a noun, meaning *plate.*
qalɛ′ Used as noun, designating kind of sgt.-at-arms in chief's council.
qö′mà vi is *blackish,* qö′mvi is *black.*
qọ̈′ya Commonly applied to the dawn.
se·′yŋaq (rather than -naq).
si′ka Also *sour.*
se′yan (rather than si-).
so′nwàyo Used by a man when he refers to a girl.
ta′levi (M) Also *slippery.*
ta′skàvi Refers often to corn which is ready to husk and to light-complected Hopis.
wiya′qti Word unknown.
ya·′sayòq Woman's word.

NUMERALS

lö·′ye In M. one hears also -yim.
pa′hiw This form not used here; rather pa′yom.
na·′lö (as variant -ye added); *fifty* is often said as *two twentys plus ten.*

TEMPORALS, ADVERBS CONJUNCTIONS, INTERJECTIONS

pa′sʔàŋʷe *time of preparing the fields, both before and after planting.*
nịq Often equivalent to our *or, but.*
pa′van Also *someone regarded as better than the next person.*
pa·′sat First vowel lengthened in M.
pẹ? Often *now.*
sɛ·′laq Usually in M. sɛ·′ɛlhaq.
ʔi′ʔic The word used in M. is ka′qtayìʔi.

pi·'hî· In M. pi·'yî·
sòniwây M. sòniwâ and sònwây.

VERBS

?ahyę'ke Unknown in this form in M. which says ?a'ŋkeyekè?e.

?e'na M. ?e'?ena.

?ena'ŋʷà? yta Also of horses, inanimate objects, desire for certain action, etc.

?o'vèkna Also of corn roasting and the like.

?ŏ'qalà Used to praise a man who has done something esteemed by his kin- or townfolk.

ce'Yti more *sneer* than *laugh* in M.

co·'na *having a good time without regard for other people.* But does not seem to have an absolutely definite pejorative sense.

ha'rèkna Also *scratches it.*

hayę'ŋna *reaching toward but not touching.*

hį'Nti Also *what happened?*

ho'ròkena (rather than -kna).

hö·'na Also *wipe.*

la'qta (rather than-qna).

lɛ''ta Also *bolts it, locks it.*

löhŏ'hötà *fruit or stones falling down,* NOT *rain.*

me'?a Also *nails it.*

me''ci This verb and its cognates are said to carry with them always the connotation *with the teeth.*

na·'kʷàvena (instead of -vna).

na·'ni Sometimes *chuckle*; sometimes *sneer.*

na·'seŋʷena (instead of -ŋna).

pa''ci Also *water falls over a damn or spillway.*

pehį'Nti (rather than pa-); mainly *feels faint.*

pa·''ta (instead of pa·'ta).

pa·'ya'la (instead of -yɛ-).

pe're Also *blooms.*

qa'?ena (rather than qɛ·'na).

re'vta Unknown in M.

se'lawe (instead of sį-).

[After this paper was set up in page proof, an overlooked reference to other recordings of Hopi words turned up, and is given here in the only available space.]

Kenneth MacLeish, Notes on Hopi Belt-Weaving of Moenkopi, AA 42.291 (1940), and Notes on Folk Medicine in the Hopi Village of Moenkopi, JAF 56.62-68 (1943).

SPANISH LOANWORDS IN HOPI; A PRELIMINARY CHECKLIST

FREDERICK J. DOCKSTADER
DARTMOUTH COLLEGE

0. Acknowledgments
1. Scope
2. The sample
3. The orthography
4. The list of borrowings
5. Other loans

0. The material presented in this article has been extracted from a manuscript Hopi Dictionary now in preparation. In compiling this dictionary, I have drawn heavily from the unpublished work of B. L. Whorf (loaned by Clyde Kluckhohn), the glossary by Whorf and A. M. Stephen,[1] Whiting's ethnobotanical study,[2] the unpublished Moencopi notes of Kenneth MacLeish, and the vocabulary studies of Clyde Kluckhohn, plus my own field work. I wish to acknowledge my indebtedness to Kluckhohn and MacLeish for their generosity.

This is in no sense a complete list of such loanwords; it is a first sifting, and much more analysis, as well as additional vocabulary and text material, will be needed before any 'complete' list can be prepared.[3]

1.0. The Hopis are similar to many other Southwestern peoples in their use of loanwords. They tend to use them occasionally in everyday speech (especially with non-Hopis), but rarely do on formal or ceremonial occasions. There is much the same tendency to 'purify' speech whenever attention is called to the fact of loanword usage.[4]

1.1. The degree of Hopi-Spanish bilingualism varies greatly among the Hopis. Some know Spanish fairly well, and use it as the situation requires; for the most part, however, Spanish is not well known. Its use is rather restricted, and many of the more conservative individuals strongly condemn Spanish in conversation. However, there are a few stock words and phrases which are widely understood, and some of these have become so much a part of the language that they are not readily recognized by the Hopis as being of Spanish origin.

1.2. The actual amount of Spanish contact with the Hopis took place under circumstances not common to most of the Pueblo tribes, and this has had an effect which this list would seem to reveal. Briefly, after the Spanish entry in 1540, there was no real 'occupation' of Hopiland until about 1630, when the first church was built at Awátovi. For the next fifty years, Spanish must have been a familiar tongue to the much-harangued Hopis, who zealously turned the rascals out when the Pueblo Revolt erupted in the Southwest. There was no significant contact with Spanish or Mexican people from that time on, insofar as linguistic influence is concerned. Therefore, it would seem that most of the Spanish loanwords which came to the Hopis directly from Spaniards (or Mexicans) date either from the period of longest Spanish residence,

[1] E. C. Parsons [ed.], Hopi Journal of Alexander M. Stephen, CU 23.1198–1326 (1936).

[2] A. F. Whiting, Ethnobotany of the Hopi (Museum of Northern Arizona) Bull. 15 (1939).

[3] This article was in part suggested by two recent articles: G. I. Trager, Spanish and English Loanwords in Taos, IJAL 10.144–158 (1944), and M. Titiev, Suggestions for the Further Study of Hopi, IJAL 12.89–91 (1946). I wish to acknowledge my obligation to both authors.

[4] This is very similar to the situation at Taos as mentioned by Trager, op. cit., p. 144; however, the Hopis are not as fanatically reticent in this respect.

1630–1680, or else from the later sporadic and short-lived contacts between members of the two groups.

1.3. A more important avenue by which these loanwords entered was undoubtedly Indians of the neighboring area. Following the Pueblo Revolt of 1680, a large number of Río Grande tribesmen fled to the Hopi villages to escape Spanish vengeance, and remained there until about 1745. Obviously this allowed sufficient time for the introduction of many Spanish terms under the more favorable guise of 'Indian' speech. After the Spanish reconquered the Río Grande area, Spanish became a familiar tongue to most of the Indians in that area. Since there was frequent intercourse between the Hopis and these somewhat Hispanicised Indians, it follows that ample opportunity would be afforded to learn new terms. I suspect that most Spanish loanwords still in use by the Hopis fall in this category.

1.4. It will be noted that for the most part, the Hopi version has retained a fairly consistent similarity to the Spanish original. This would support the argument that these words are largely of recent adoption, or else that they have undergone such extreme change that their origin is unrecognizable.

2.0. This checklist includes some 66 words which have been culled from a vocabulary of approximately 4500 terms. This suggests that the proportion of accepted Spanish terms in general Hopi use is quite small. A large majority of these words are nouns relating to objects of foreign introduction; there are very few such words which refer to things native to the Hopi area—and it is probable that these include some which are linguistic coincidence. In a few instances, it may be that the native name became displaced by the Spanish term, and is now lost.

I have tried to be most conservative in selecting words for this listing. Even so, there are some whose Spanish origin seems quite dubious—these I have indicated by (?), and list them only for speculative purposes.

3.0. Since the primary purpose of the dictionary mentioned above is ethnological rather than linguistic, this checklist will reflect that fact. I make no pretence at being a trained linguist, and have employed an orthography sufficiently simple to enable the non-Hopi to reasonably duplicate the sounds so as to be intelligible to the native.

3.1. For the most part I have followed Whorf's key,[5] with the following exceptions: all vowels are as in "Continental" usage; ü approximates [eu] of French fleuve; ch in all cases is as English church.

3.2. Dialect differences have not been indicated. This list largely reflects Second Mesa usage, with any major variants from other Mesas listed second. In many cases the precise rendition of a word will vary not only with the several Mesas, but even with speakers from different villages on the same Mesa. This whole problem of "Hopi dialects" is largely unexplored territory, and remains a subject of strong debate.

4.0.

ʔáʔa *jaybird*, /gayo. Possibly, /gaya *magpie*.

ʔaʔáy *listen!* exclamation, /oye!

ʔáʔane *quickly!* exclamation, /ándale!

ʔaíya'ada *work-party*, /ayuda, ayudante.

ʔá·ne *ant*, /anay [a species similar to the termite]

ʔá·sa *mustard*, /mostaza (?). [Historically, Asa Clan is one of the earliest, and figures prominently in mythology. This suggests coincidence, rather than derivation].

ʔasáfrani *safflower*, /azafrán.

ʔayám *there, over there*, /allá.

ʔaya'qá *there* [closer], /acá.

bánda *hair turban*, /banda.

béso *dollar*, /peso [possibly via the Navajo].

chácüs *syphilis*, /chancro

[5] Parsons, op. cit., pp. 1199–1201.

cháko *little, small*, /chico (?).
héya, héyi *trade, barter*, /canjear (?)
ho?ápe, hoápü *basket, sack*, /hopa
kané·lo *sheep*, /carnero
kóhe, kápe *coffee*, /café [possibly via Navajo].
kapíri, káprizi *goat*, /cabrío, cabrerizo
karé·ta *cart, wagon*, /carreta
kasá·vü *casaba*, /cazabe
ká·sta *enough*, /basta
kástila *Spaniard* [and *Mexican*], /castellano
kawaíyo *horse*, /caballo
kawá'yo *watermelon*, /caballo [Hopis refer to the melon as smelling like a sweaty horse, hence the term.]
kípo *war cry*, /grito (?)
kókyaŋwü, kókyaŋa *spider*, /araña
komántsi *Comanche*, /comanche
korása, kolása *leather*, /cuero (?)
kowáko *chicken*, /guaco or perhaps, /cloqueo [although onomatopoeia is uncommon in Hopi.]
koyóŋo *turkey*,/gallina
kyá·ro, kyar *parrot*, /loro or perhaps, /papaguayo.
lánsa *spear*, /lanza
léŋi *tongue*, /lengua (?)
limón *lemon*, /limón
limónigi?i *lemonade*, /limonada
máta, má·da *mano stone*, /mano (?)
mansána *apple*, /manzana
meló·ni *melon*, /melón
mó·la *mule*, /mula
mó·ro *donkey*, /burro
müsa *cat*, /miza
müyi *mole*, /mola (?)
ŋerésana, ŋé·re *rubber*, /gomorresina, seringa
né·va, nüva *snow*, /nieve (?)
ŋö·la *ring*, /anillo, argolla
?ö'qala, óqala *greetings*, an exclamation, /ojalá
?ö·va *grape*, /uva
pása?a *finished, done*, /pasada
pavóya *sprinkle with pollen*, /polvorear (?)
pénta *writes, paints*, /pintar
pichóti *pig*, /cochino (?)
pí·nto [mori] *pinto bean*, /pinto
rá·na *little frog*, /rana
silántro *coriander*, /cilantro
sio'úyi *onion*, /cebolla [Literally, "Zuñi plant"; referring to its Zuñi origin. Yet it is of non-Indian introduction, as Whiting has shown,[6] and the form suggests Spanish derivation.]
siyára, siárra *saw, sword*, /sierra
ta·váco *tobacco* [Spanish style], /tabaco
tokóri *a small owl*, /tecolote
tomáti *tomato*, /tomate
tsíli, chíli *chili*, /chili
tsómo, chómo *mount, hill*, /cima, loma
wá·kasi, wakási *cow*, /vaca
wá·ya?a *go away!* exclamation, /vaya!
yaláha *micaceous hematite*, /hierro (?)
yó·wi *rain*, /lluvia (?)
yúŋya *cactus fruit*, /tuna

5.0. There has been no attempt to consider compounds, nor those Hopi terms descriptive of objects [loan-translations]; these are of a different character from those for which this article is intended.

5.1. More extensive vocabulary and text material will certainly reveal more terms of Spanish origin. However, I believe the proportion of Spanish:Hopi will not be found to undergo marked change. It is very doubtful whether many added loanwords will be introduced in the future—Spanish-speaking people are not common in the Hopi area, nor is their speech particularly popular. English, of course, will be responsible for an increasing number of loanwords until such time as full acculturation takes place. In view of Hopi history and conservatism, that time seems far distant.

[6] Whiting, op. cit., p. 9, 12.

HIDATSA II: AFFIXES*

FLORENCE M. ROBINETT
COLORADO SPRINGS

2.0. Introduction
2.1. Affix list
2.2. Affix sequences
2.3.1. Sub-sets of person markers
2.3.2. Person marking paradigms
2.3.3. Position of paradigms
2.3.4. Paradigm overlap

2.0. Numbers are here assigned to affixes as an index for the morphemes in question and also as an indicator of their positions. The affixes are listed by decade classes, such that the members of each class occur in the same position in respect to other classes of morphemes. The affixes within a decade class are mutually exclusive. In any sequence, affixes occur in the numerical order of their decade class. (Prefixes of Class 10 precede those of Class 20, which precede those of Class 30, and so on.) Since many of the classes of Hidatsa affixes—many of which contain only one member—may occur either before or after certain other classes or themselves, it was necessary to assign more than one decade number to some of the classes. Thus, prefix i·- *instrumentive*, which may precede prefixes of all other classes and may follow prefixes of only four classes, was assigned to decade Class 10–60. Members of Class 10–60 precede or follow members of Classes 20–70, 30–80, 40–90–110, and 50. Members of Class 20–70 precede or follow members of Class 10–60 and Class 30–80, and so on. For convenience, in reference only the first positional number of multiple decade numbers is given in the numbers of specific affixes, as 11 for prefix i·-, above, rather than 11–51. Thus in listing the morphemes in i·té·he *weapon*, we write:
11 + stem *die* + 1021.

Decade numbers below 1000 are assigned to classes of prefixes, those above 1000 to classes of suffixes.

2.1. The affixes are:

10–60.
- 11. i·- instrumentive

20–70.
- 21. wa·- indefinite (subject, object, possessor)

30–80.
- 31. aru- ~ oru- future, agentive, *sphere of, place where*
- 32. aku- agentive

40–90–110–1030–1110–1180
- 41. wi-(.1, .8), wih-(.1), wi·-(.2), wa-(.3, .4, .5, .6), wah-(.6), w-(.7), -wa·(.9), -wi(.10) first person
- 42. ri-(.1, .3, .4, .8), rih-(.1), ri·-(.2), ra-(.5, .6), rah-(.6), r-(.7), -ra·(.9), -ri(.10) second person
- 43. i-(.1, .3), ih-(.1), a-(.5), zero-(.2, .4, .6, .7, .8), -zero(.9, .10) third person

50.
- 51. ta-(.1), t-(.2) alienable

100–120.
- 101. ki- suus, reflexive, *become*

130.
- 131. k- completive

140.
- 141. ru- *by hand*
- 142. ra- *by teeth*
- 143. ara- (.1) ,-ʔa- (.2) *by foot*
- 144. pa- *by force, with stick*
- 145. ha- *with knife*
- 146. ka- *with hammar, fist*

*For Hidatsa I: Morphophonemics, see IJAL 21.1–7 (1955); the examples again follow in the order: HIDATSA ENGLISH (MORPHEME FORMULA).

1000–1210.
1001. -hta *along, toward, in the position of,* implementive
1010–1140.
1011. -a·cI (.1) -racI (.2) approximative
1012. -ra· approximative
1020.
1021. -he· (.1), -hi (.2), -zero (.3) causative, *use for /as*
1040.
1041. kša iterative
1050.
1051. -wi quantitative
1060.
1061. -ria reflexive
1070–1130.
1071. -ʔi· durative, iterative
1072. -ahi momentaneous
1090–1150
1091. -tha negative
1100.
1101. -hke· (.1), -hki (.2) causative, diminutive, imitative
1120.
1121. -hti desiderative
1170.
1171. -ha in reference to time, place
1190.
1191. -hi (.1), -zero (.2) future, probability, ability
1200.
1201. -aʔa ~ oʔo (.1), -ru (.2) plural
1202. -apa plural (group)
1220.
1221. -he *that one here*
1222. -ri objectifier, demonstrative
1223. -ka *that one*
1230.
1231. -kua (.1), -hkua (.2) *in, into, that one*
1232. -kuha· (.1), -hkuha· (.2) *from*
1240.
1241. -ši in reference to time
1250.
1251. -ru inclusive (time or place), *if*
1252. -to· indefinite, *or*
1253. -šta· ~ -št habitual agent, past definite
1260.
2161. -ware. quotative
1270–1280.
1271. -c predicative, utterance final
1272. -š substantive, *named*
1273. -wa *as, when, at*
1274. -ša concessive
1275. -še distributive
1276. -hiri *because*
1277. -rahe reportative
1278. -ak (.1), -k (.2) predicative, non-utterance final
1279. -ʔrepeat-final-mora interrogative
1281. -ara imperative plural
1282. -zero ~ -h ~ -ka (.1) -a ~ -ah ~ aka (.2) imperative singular
1283. -ški emphatic

2.2. We now discuss affix sequences by decade classes.

Prefix Class 10–60

Prefix 11, i·- instrumentive, occurs word initially or after prefixes of Classes 20–70, 30–80, 40–90–110 and 50. Sequences in which prefix 11 occur are of the following types.[1]

11 + S ± 1021: i·pakíši *towel* (11 + S), i·té·he *weapon* (11 + S + 1021).

11 ± 21 + 40 ± 140 + S + suffixes: i·ʔú·ʔa·k *he climbed by means of (it)* (11 + 43 + S + 1278), i·wa·ʔi·káhe·thac *by it he blinds him* (11 + 21 + 43 + S + 1021 + 43 + 1091 + 1271), i·wa·paxúxhic *he broke it (something) off* (11 + 21 + 43 + 144 + S + 1271), i·rahcíc *he bit it off* (11 + 43 + 142 + S + 1271).

± (40 + 50) + 21 + 11 ± 131 + S: wa·ʔi·ʔí·he *paint* (21 + 11 + S), watawa·ʔi·ʔí·he *my paint* (41 + 51 + 21 + 11 + S), wa·ʔi·kí·ka *window* (21 + 11 +

[1] Only the first number of multiple decade numbers is used here, as 40 rather than 40 90–110–1030–1110–1180 in the second formula.

131 + S), itawa·ʔi·kí·ka *his window* (43 + 51 + 21 + 11 + 131 + S).

21 ± 31 + 11 + 40 + S: wa·i·ré *what he's going for* (21 + 11 + 43 + S), wa·ʔoruʔi·wihapá·ci *what I'd be glad about* (21 + 31 + 11 + 41 + S).

11 + 40 + 101 + S + 1101 + 40 + other suffixes: i·wikicé·šahke·c *I'm greedy* (= *turned into a wolf with it*) (11 + 41 + 101 + S + 1101 + 43 + 1271), i·wikixá·pihke·thac *I'm not getting thin* (11 + 41 + 101 + S + 1101 + 43 + 1091 + 1271).

Prefix Class 20–70

Prefix 21, wa·- *indefinite* occurs word initially or after prefixes of Classes 10–60, 30–80, 40–90–110, and 50. Sequences in which prefix 21 occurs are of the following types.

± (40 + 50) + 21 ± 30 + S ± 1101: wa·ʔapá·ri *weed* (21 + S), watawa·ʔapá·ri *my weed* (41 + 51 + 21 + S), wa·ʔaruhí·ši *something red* (21 + 31 + S), wataʔwa·ʔ-arupóhoruwi *my bullet* (41 + 51 + 21 + 31 + S), wa·ʔakupóhoruwihke *apple* (21 + 32 + S + 1101), itawa·ʔakupóhoruwihke *his apple* (43 + 51 + 21 + 32 + S + 1101).

± (40 + 50) + 21 + 11 ± 131 + S: for examples see prefix 11, above.

21 + 31 + 21 + S: wa·ʔoruʔwa·pá·hi *the singing* (21 + 31 + 21 + S).

32 ± (40 + 50) + 21 + S ± 1021: akuwa·ʔáši *hunter* (32 + 21 + S), akuwa·-kírikihe *blacksmith* (32 + 21 + S + 1021), akuʔitawa·ʔéʔe *owner* (32 + 43 + 51 + 21 + S).

21 ± 31 ± 11 + 40 + S: wa·ʔaruwahé *what I'm doing* (21 + 31 + 41 + S); for other examples see prefix 11, above.

± 11 + 21 + 40 ± 140 + S + suffixes: wapá·hic *he sings* (21 + 43 + S + 1271), wa·paxpíc *he bumped it* (21 + 43 + 144 + S + 1271); for other examples see prefix 11, above.

((40 + 21) ~ (21 + 40)) + 40 + S + suffixes: wa·ri·wahkiwáxhuc (21 + 42 + 41 + S + 1271) ~ ri·wa·wahkiwáxhuc (42 + 21 + 41 + S + 1271) *I ask you something.*

40 + 21 + 31 + 101 + S + 1021 + other suffixes: wiʔwa·ʔarukiréche·racic *she ignored me* (41 + 21 + 31 + 101 + S + 1021 + 43 + 1011 + 1271).

31 + 21 + 40 + S + suffixes: aruwa·-wahé·c *I'll work* (31 + 21 + 41 + S + 1271).

40 + 32 + 21 + S + suffixes: riʔakuwa·-híric *you're a worker* (42 + 32 + 21 + S + 1271).

Prefix Class 30–80

Prefixes of Class 30–80 (31. aru- ~ oru- future, agentive and 32. aku- agentive) occur word initially or after prefixes of Classes 20–70, 40–90–110 and 50.

The variation between the two allomorphs of prefix 31 (aru-, oru-) is free.

Sequences in which prefix 31 occurs are of the following types.

± (40 ± 50) ± 21 + 31 + S ± suffixes: aruʔácha *the near one* (31 + S), aruháckihta *lengthwise* (31 + S + 1001), wi·ʔaruráxpi *my skin* (41 + 31 + S), wi·ʔaruʔáto ʔo *my sores* (41 + 31 + S + 1201), itaʔaruwía *his wife* (43 + 51 + 31 + S); for additional examples see prefix 21, above.

31 + 40 + 50 + S: aruʔitawa·píša *his greatgrandchild* (31 + 41 + 51 + S).

31 ± (40 or 21) + 40 ± 140 + S + suffixes: aruwiʔákiš *where I joined* (31 + 41 + S + 1272), aruwa·rí?ic *I'll shoot it* (31 + 43 + 41 + S + 1271), aruruxúxhic *he'll break it* (31 + 43 + 43 + 141 + S + 1271); for additional examples see prefix 21, above.

+ 40 ± 21 + 31 ± 101 + S + 1021 + other suffixes: arukó·wihe·thaʔac *they will never stop* (43 + 31 + S + 1021 + 1091 + 1201 + 1271); for additional examples see prefix 21, above.

21 + 31 ± 11 + 40 + S and 21 + 31 + 21 + S. For examples of these see prefix 21, above.

Prefix 32, aku- agentive, occurs in sequences of the following types.

32 + S ± suffixes: akušíˑpi *a thick place, thicket* (32 + S), akuʔóˑthithaʔaš *the ones that are not ripe* (32 + S + 1091 + 1201 + 1272).

± (40 + 50) + 32 + 40 + S: akuʔiréʔe *interpreter*, (32 + 43 + S), itaʔakuʔiréʔe *his interpreter* (43 + 51 + 32 + 43 + S).

± (40 + 50) + 21 + 32 + S + 1101; 32 ± (40 + 50) + 21 + S ± 1021; and 40 + 32 + 21 + S + suffixes. For examples of these see prefix 21, above.

Class 40–90–110–1030–1110–1180

The prefix allomorphs of the members of Class 40–90–110–1030–1110–1180 (41 wi- (.1, .8), wih- (.1), wiˑ- (.2), wa- (.3, .4, .5, .6), wah- (.6), w- (.7) first person; 42 ri- (.1, .3, .4, .8), rih- (.1), riˑ- (.2), ra- (.5, .6), rah- (.6), r- (.7) second person; 43 i- (.1, .3), ih- (.1), a- (.5), zero- (.2, .4, .6, .7, .8) third person) occur word initially or preceded by prefixes of Classes 10–60, 20–70, 30–80, and 100–120.

The suffix allomorphs (41 -waˑ (.9), -wi (.10) first person; 42 -raˑ (.9), -ri (.10) second person; 43 -zero (.9, .10) third person) occur word medially; alternants point 9 occur immediately preceded by suffix 1021 or 1101 and followed by another suffix or suffixes; alternants point 10 occur immediately followed by suffix 1191 and preceded by a stem or a suffix other than 1021 or 1101.

The distribution of particular allomorphs is determined by the morphemic composition of the sequence in which they occur—specifically, by the theme in the sequence or by the occurrence of prefix 51 or suffixes 1021, 1101, or 1191 in the sequence.

The allomorphs are numbered on the basis of their mutual membership in sets which occur in particular sequences. For their alignment into sets and paradigms and the specific decade positions of each paradigm see **2.3.**

The variation between the two point 6 allomorphs of 41 and 42 and the two point 1 allomorphs of 41, 42 and 43 is phonemically determined; the alternants ending in /h/ (wih-, rih-, ih- (.1); wah-, rah- (.6)) occur before /k/ when not followed by /h/; the alternants ending in vowels (wi-, ri-, i- (.1); wa-, ra- (.6)) occur elsewhere.

Sequences in which Class 40–90–110–1030–1110–1180 occurs are of the following types.

± other prefixes + 40 ± 50 ± other prefixes ± 131 + S ± suffixes: wiʔí *my mouth* (41.1 + S), wiʔíc *it's my mouth* (41.1 + S + 1271), wataʔáwa *my land* (41.3 + 51 + S), wataʔáwoʔo *our land* (41.3 + 51 + S + 1201), ritawaˑʔiˑkíˑka *your window* (42.3 + 51 + 21 + 11 + 131 + S); for other examples of sequences including prefixes 11, 21, 31 and 32 see the examples listed under each of these above.

± other prefixes ± 40 ± 21 + 40 ± other prefixes except 51 ± 140 + S ± suffixes: wíˑrac *I yawn* (41.1 + S + 1271), waʔaxúac *I hide it* (43.2 + 41.5 + S + 1271), kikhíc *he drank up his own* (43.6 + 101 + 131 + S + 1271), kiruškíwareˑc *he pulled out his own, they say* (43.6 + 101 + 141 + S + 1261 + 1271), waʔašúˑkic *I wiped it with my feet* (43.2 + 41.5 + 143 + S + 1271); for examples of sequences which include prefixes 11, 21, 31 and 32 see the examples listed under each of these above.

± 31 ± 11 + (40 ± 21) or (21 ± 40) or 101 ± other prefixes except 51 ± 140 + S ± 1001 + (1021 + 40 ± other suffixes ± (1101 + 40)) or (1101 + 40) + other suffixes: cákheˑc *he made him mourn* (43.2 + S + 1021 + 43.9 + 1271), arucákheˑc *he will make him mourn* (31 + 43.2 + S + 1021 + 43.9 + 1271), kiréˑcheˑc *he uses up his own* (101 + S + 1021 + 43.9 + 1271), arukiréˑcheˑc *he will use up his own* (31 + 101 + S + 1021 + 43.9 + 1271), paxúxhiheˑc *he made him break it off* (43.2 + 144 + S + 1021 + 43.9 + 1271), wiˑkárawiheˑracic *he makes me remember a little bit* (41.2 + S + 1021 + 43.9 + 1011 + 1271), kárawiheˑracihkeˑc *he makes him remember it a little* (43.2 + S + 1021 + 43.9 + 1011 + 1101 + 43.9 + 1271), iˑwikixáˑpihkeˑc *I'm getting thin* (11 +

41.8 + 101 + S + 1101 + 43.9 + 1271); for examples of sequences which include prefixes 11, 21 and 32 see the examples listed under each of these, above.

± (40 ± 21) or (21 ± 40) or 101 ± other prefixes except 51 + S ± (1021 + 40) ± other suffixes ± (1101 + 40) + (40 + 1191) + other suffixes: cixíhic *he'll jump* (S + 43.10 + 1191 + 1271), axúawic *I'll hide it* (43.2 + S + 41.10 + 1191 + 1271), kiré·chiwa·wic *I'll use up my own* (101 + S + 1021 + 41.9 + 41.10 + 1191 + 1271), cixíhkiwa·wiha?ac *we'll make him jump* (43.2 + S + 1101 + 41.9 + 41.10 + 1191 + 1201 + 1271), wa·hírihic *he'll work* (21 + S + 43.10 + 1191 + 1271), kárawihe·racihke·hic *he'll make him remember it a little* (43.2 + S + 1021 + 43.9 + 1011 + 1101 + 43.9 + 43.10 + 1191 + 1271).

40 + 40 + 50 + S + 1021 + 40 + other suffixes: ri·watacákhiwa·c *I like you* (= *I make you my good*) (42.2 + 41.3 + 51 + S + 1021 + 41.9 + 1271), wi·?itacákha·?ac *he likes us* (41.2 + 43.3 + 51 + S + 1021 + 43.9 + 1201 + 1271).

40 + 50 + 32 + 40 + S: ita?aku?iré?e *his interpreter* (43.3 + 51 + 32 + 43.3 + S).

Prefix Class 50

Prefix 51, ta- (.1), t- (.2) alienable, occurs word medially, immediately preceded by any point three allomorph of Class 40 and followed by another prefix or a stem.

The variation between the two allomorphs of prefix 51 is determined by the theme in a given sequence; ta- occurs in sequence with most themes which occur in sequence with prefix 51; t- occurs with a few such themes. (The themes in our sample which occur with t- are: ú·waki *bed, blanket* úaxi *shirt, coat* and xúpa·ri *medicine*; i·púxi *rattle* occurs with either allomorph of 51.)

Sequences in which prefix 51 occurs are of the following types.

± other prefixes + 40 + 51 ± other prefixes ± 131 + S ± suffixes: wata?íri *my blood* (41.3 + 51 + S), ita?ó·pi *his tobacco* (43.3 + 51 + S), aru?itawa·píša *his greatgrandchild* (31 + 43.3 + 51 + S), aku?itawa·?é?e *owner* (32 + 43.3 + 51 + 21 + S), watawa·?i·kí·ka *my window* (41.3 + 51 + 21 + 11 + 131 + S), ita?i·té·he *his weapon* (43.3 + 51 + S + 1021), rita?áwa?ac *it's your (plural) land* (42.3 + 51 + S + 1201 + 1271).

40 + 40 + 51 + S + 1021 + 40 + other suffixes: ri·watacákhiwa·c *I like you* (42.2 + 41.3 + 51 + S + 1021 + 41.9 + 1271).

40 + 50 + 32 + 40 + S: ita?aku?iré?e *his interpreter* (43.3 + 51 + 32 + 43.3 + S).

Prefix Class 100–120

Prefix 101, ki- suus, reflexive, *become*, occurs word initially or after prefixes of Classes 10–60, 20–70, 30–80 and 40–91–110.

Sequences in which prefix 101 occurs are of the following types.

± other prefixes + 101 ± 131 ± 140 + S + suffixes: kiré·che·c *he uses up his own* (101 + S + 1021 + 43.9 + 1271), kikhíc *he drank up his own* (43.6 + 101 + 131 + S + 1271), ihkirušú·kic *he cleaned himself* (43.1 + 101 + 141 + S + 1271), kitúaru *when it turns green* (43.8 + 101 + S + 1251), i·wikihíšihke·c *I'm blushing* (= *become reddened by it*) (11 + 41.8 + 101 + S + 1101 + 43.9 + 1271), arukitó?ohihke·c *he'll turn blue* (31 + 43.8 + 101 + S + 1101 + 43.9 + 1271), wahkikí·ric *I looked for it* (43.2 + 41.6 + 101 + S + 1271).

(101 + 40) or (± 101 + 40 + 101) + S + suffixes: kiwirahcá·c *I'm beginning to get careful* (101 + 41.1 + S + 1271), kiwi·hí·rahpic *I'm getting into a hard time* (101 + 41.2 + S + 1271), kiwi·haxíša·cic *I'm wet, but not very* (101 + 41.2 + S + 1011 + 1271), kiwi·xí?e·c (101 + 41.2 + S + 1271) or wi·kixí?e·c (41.2 + 101 + S + 1271) or kiwi·kixí?e·c (101 + 41.2 + 101 + S + 1271) *I'm getting old.*

Prefix Class 130

Prefix 131, k- completive, occurs word initially (in the sense of being the first

phonemically constituted element of the word) only when preceded by a zero alternant of prefix 43, third person; elsewhere, it occurs word medially, preceded by prefixes of Classes 10–60, 20–70, 30–80, 40–90–110, 50, and 100–120 and followed by prefixes of Class 140 or stems.

Sequences in which prefix 131 occurs are of the following types.

40 + 40 or 101 + 131 ± 140 + S + suffixes: kikhíc *he drank up his own* (43.6 + 101 + 131 + S + 1271), wahkaxúac *I hide it away* (43.2 + 41.6 + S + 1271), wahkaraxúxhic *I broke it by stepping on it* (43.2 + 41.6 + 143 + S + 1271).

Prefix Class 140

Prefixes of Class 140 (instrumentals 141. ru- *by hand*, 142. ra- *by teeth*, 143. ara- (.1), -ʔa- (.2) *by foot*, 144. pa- *by force, with stick*, 145. ha- *with knife*, 146. ka- *with hammer, fist*) occur word initially only when preceded by a zero alternant of prefix 43, third person, elsewhere they occur word medially, preceded by prefixes of Classes 10–60, 20–70, 30–80, 40–90–110, 100–120, and 131 and followed by stems.

The variation between the two alternants of prefix 143 is morphemically determined; -ʔa- occurs when 143 is immediately preceded by first or second person actor prefixes (41.6 or 42.6); ara- occurs in other sequences.

Prefixes of Class 140 occur with only about seventy stems in our sample, each of which occurs with only certain members of Class 140. (These stems and the instrumental prefixes with which each occurs are listed in **3.**)

Sequences in which prefixes of Class 140 occur are of the type: ± other prefixes + 40 ± other prefixes + 140 + S + suffixes: rušú·kic *he wiped it with his hands, he washed it* (43.2 + 43.6 + 141 + S + 1271), rahcakic *he bit it off* (43.2 + 43.6 + 142 + S + 1271), arašú·kic *he wiped it with his foot* (43.2 + 43.6 + 143.1 + S + 1271), waʔašú·kic *I wiped it with my foot* (43.2 + 41.6 + 143.2 + S + 1271), paxpíc *he pushed it off* (43.2 + 43.6 + 144 + S + 1271), hašíšhic *he cuts it by putting point of knife in* (43.2 + 43.6 + 145 + S + 1271), kaxpíc *he breaks it by pounding* (43.2 + 43.6 + 146 + S + 1271); for examples of sequences which include prefixes of Class 140 and prefixes 11, 21, 31, 40, 101, and 131 see the examples listed under each of these above.

Suffix Class 1000–1210

Suffix 1001, -hta *along, toward, in the position of*, implementive, occurs word finally or followed by suffixes of Classes 1020, 40–90–110–1030–1110–1180, 1200, 1220, and 1270. Sequences in which suffix 1001 occurs are of the following types.

± 31 or 40 + S + 1001 ± 1201 or 1222: awašíhta *to the cellar* (S + 1001), wíʔihta *with a stone* (S + 1001), réhta *along the edge* (S + 1001), wacíhtaʔa *with my feet* (41 + S + 1001 + 1201), íhá·htari *in the others* (S + 1001 + 1222), aruháckihta *lengthwise* (31 + S + 1001).

40 + S + 1201 + 1001: wi·cáhaʔahta *by my canes* (41.2 + S + 1201 + 1001).

S + S + 1001 + 1021 + 40 + 1271: í·awahuhtahe·c *he turns fur inside* (S + S + 1001 + 1021 + 43.9 + 1271).

Suffix Class 1010–1140

Suffix Class 1010–1140 (1011-a·cI (.1), -racI (.2) approximative; 1012) -ra· approximative) occurs word medially preceded by a stem or a suffix of Class 40 (preceded by suffix 1021), 1040, 1090, 1100, or 1120, and followed by suffixes of Classes 1020 (plus 40), 1070, 1100 (plus 40) or 1200, plus 1270.

The variation between the two alternants of suffix 1011 is for most sequences phonemically determined— -a·ci occurs when 1011 is preceded by a consonant or a single mora vowel, -raci occurs when 1011 is preceded by a long vowel; but, in sequence

with a few stems either may occur, with a slight shift in the informant's translation, as: irá?a·cic *he kind of talks* (43 + S + a·ci + 1271), iré?eracic *he talks a little* (43 + S + raci + 1271).

Sequences in which suffix 1011 occurs are of the following types.

40 + S ± 1091 or 1121 + 1011 + 1270: tó·pa·cic *it's around four* (43 + S + 1011.1 + 1271), wa·ré·tha·cic *I kind of didn't go there* (41 + S + 1091 + 1011.1 + 1271). irá?a·cišha *he's kind of talking, but—* (43 + S + 1011.1 + 1274)[2], iré?ehta·cic *he kind of likes to talk* (43 + S + 1121 + 1011.1 + 1271).

40 ± S + S ± 1041 + 1011 + 1071 + 1271: waxúara?a·ci?i·c *I'm always a little sick* (41 + S + S + 1011.1 + 1071 + 1271), iré?ekša·ci?i·c *he always keeps on kind of talking* (43 + S + 1041 + 1011.1 + 1071 + 1271).

40 + S + 1021 + 40 + 1011 ± (1101 + 40) or 1071 + 1271: wi·kíthe·racic *he cheated me* (41 + S + 1021 + 43 + 1011.2 + 1271), kárawiwa·racihkiwa·c *I make him remember it a little* (43 + S + 1021 + 41 + 1011.2 + 1101 + 41 + 1271), aré?ehe·raci?i·c *he gets kind of mad every now and then* (43 + S + 1021 + 43 + 1011.2 + 1071 + 1271).

40 + 101 + S + 1011 + 1021 + 40 + 1271: kihaxíša·chiwa·c *I make him a little wet* (43 + 101 + S + 1011.1 + 1021 + 41 + 1271).

S ± 1101 + 1011 ± 1201 + 1271: atí?a·cic *it's a no good house* (*it's kind of a house*) (S + 1011.1 + 1271), atíhke·raca?ac *these small houses are no good* (S + 1101 + 1011.2 + 1201 + 1271).

21 + 31 + S + 1011 + 1272: wa·?aru?-išíaraciš *the bad one* (21 + 31 + S + 1011.2 + 1272).

Suffix 1012 occurs in sequences of the type 40 + S + 1012 + 1271, as watíra·c *I almost died* (41 + S + 1012 + 1271) (cp. watá·cic *I fainted*, 41 + S + 1011 + 1271).

[2] For the alternation of morpheme final vowels before morpheme initial /a/ see **1.2.**

Suffix Class 1020

Suffix 1021, -he· (.1), -hi (.2), -zero (.3) causative, *use for or as*, occurs word finally immediately preceded by a stem, or word medially preceded by a stem or suffix 1001 or 1011 and immediately followed by any point nine (suffix) alternant of Class 40 plus suffixes of Classes 1010–1140, 1040, 1060, 1070–1130, 1090–1150, 1100, 1190, 1200, 1250, or 1260 and 1270, or word medially preceded by a stem and followed by suffixes of Classes 1200, 1250, or 1270.

The variation between the three alternants of suffix 1021 is in part morphemically and in part phonemically determined; alternant .2 (-hi) occurs when 1021 is immediately preceded by a vowel and followed by 41.9 or 42.9 (first and second person actor); alternant .3 (-zero) occurs when 1021 is immediately preceded by a consonant and followed by 41.9 or 42.9; alternant .1 (-he·) occurs elsewhere.

Sequences in which suffix 1021 occurs are of the following types.

± 31 ± 11 + (40 ± 21) or (21 ± 40) or 101 ± other prefixes except 51 + S + 1001 or 1011 + 1021 + 40 ± other suffixes + 1270: wikáhukhe·c *he fooled me* (41.8 + S + 1021.1 + 43.9 + 1271), wikáhukhira·c *you fooled me* (41.8 + S + 1021.2 + 42.9 + 1271), táwa·c *I killed him* (43.2 + S + S + 1021.3 + 41.9 + 1271), wa·tára·?a *did you kill someone?* (21 + S + 1021.3 + 42.9 + 1279), aruwa·táhe·c *he will kill someone* (31 + 21 + S + 1021.1 + 43.9 + 1271), arukó·wihe·tha?ac *they will never stop* (31 + 43.2 + S + 1021.1 + 43.9 + 1091 + 1201 + 1271), kiré·che·c *he uses up his own* (101 + S + 1021.1 + 43.9 + 1271), kárawihe·racihke·hic *he'll make him remember it a little* (43.2 + S + 1021.1 + 43.9 + 1011 + 1101 + 43.9 + 43.10 + 1191 + 1271), aphúhkaha·k *using it as a hat* (43.8 + 43.5 + S + 1021.1 + 43.9 + 1278); for other examples see the examples given

under prefixes 11, 21, 31, 32 and 40 and suffixes 1001 and 1011 above.

± (40 + 50) + 11 or 31 + S + 1021 ± 1251: i·té·he *weapon* (11 + S + 1021.1),[3] ita?i·té·he *his weapon* (43.3 + 51 + 11 + S + 1021.1), oru?í·rihe *a loan, lease* (31 + S + 1021.1), ita?aru?apaxíhe·ru *his place for resting* (43.3 + 51 + 31 + S + 1021.1 + 1251).

32 ± 21 + S + 1021 ± 1201 or 1253 ± 1279: akutáhe *one who was killed* (32 + S + 1021.1), akutáha·?a *ones who were killed* (32 + S + 1021.1 + 1201), akutáhe·šta *the one who was killed* (32 + S + 1021.1 + 1253), akutáhe·šta?a *which one is it that was killed?* (32 + S + 1021.1 + 1253 + 1279), akuwa·kírikihe *blacksmith* (32 + 21 + S + 1021.1).

40 + 40 + 50 + S + 1021 + 40 + 1271: ri·watacakhiwa·c *I like you* (42.2 + 41.3 + 51 + S + 1021.2 + 41.9 + 1271).

S + S + 1021: atícuahe *tipi* (= *house put up conical*) (S + S + 1021.1).

Suffix Class 1040

Suffix 1041, -kša iterative, occurs word medially, preceded by a stem or suffix 1021 plus 40 and followed by suffixes of Classes 1010–1140, 1070–1130, or 1200 and 1270.

Sequences in which suffix 1041 occurs are of the following types.

40 + S ± (1021 + 40) + 1041 ± 1201 + 1271: iré?ekšac *he talks all the time* (43 + S + 1041 + 1271), watíkša?ac *we faint* (41 + S + 1041 + 1201 + 1271), aré?era·kšac *you're mad all the time* (43 + S + 1021 + 42 + 1041 + 1271), aré?ehe·kša?ac *they're mad all the time* (43 + S + 1021 + 43 + 1041 + 1201 + 1271).

40 + S + 1041 ± 1011 + 1071 + 1271: té·kša?i·c *he always faints* (43 + S + 1041 + 1071 + 1271), iré?ekša·ci?i·c *he always keeps on kind of talking* (43 + S + 1041 + 1011 + 1071 + 1271).

[3] For the variation of morpheme final long vowels with morpheme final short vowels see 1.3.

Suffix Class 1050

Suffix 1051, -wi quantitative, occurs word finally preceded by a stem, or word medially preceded by a stem and followed by suffix 1171.

The following are examples of forms containing suffix 1051: túawi *how many* (S + 1051), túawiha *how many times* (S + 1051 + 1171), šíʔawi *that many* (S + 1051), šíʔawiha *that many times* (S + 1051 + 1171).

Suffix Class 1060

Suffix 1061, -ria reflexive, occurs word medially preceded by a stem or a suffix of Class 40 (preceded by suffix 1021) and followed by suffixes of Classes 1070, 1100 (plus 40) or 1200 plus 1270.

Suffix 1061 occurs in sequences of the following types.

40 ± 101 + S + 1061 ± (1071 + 1101 + 40) ± 1201 + 1271: wi·cákiriac *I am fussy (good to/for me)* (41 + S + 1061 + 1271), wihkicákiria?i·hkiwa·c *I always make myself fussy* (41 + 101 + S + 1061 + 1071 + 1101 + 41 + 1271), wihkicákiria?i·hkiwa·?ac *we always make ourselves fussy* (41 + 101 + S + 1061 + 1071 + 1101 + 41 + 1201 + 1271), kí?iria?a?ac *they ride horseback* (43 + S + 1061 + 1201 + 1271).

40 + S + 1021 + 43 + 1061 + 1270: wi·híše·riac *I made myself red* (41 + S + 1021 + 43 + 1061 + 1271).

Suffix Class 1070–1130

Suffix Class 1070–1130, 1071. -ʔi·- durative, iterative, 1072. -ahi- momentaneous, occurs word medially preceded by a stem or by suffixes of Classes 1010–1140, 1020 plus 40, 1040, 1060, 1090, or 1100 plus 40, and followed by suffixes of Classes 1100 plus 40, 40 plus 1190, 1200, 1250, 1260, or 1270.

Suffix 1071 occurs in sequences of the following types.

40 ± 101 + S ± 1041 ± 1011 + 1071 ± 1201 or 1261 + 1271 or 1283: círi?i·c *it's always yellow* (43 + S + 1071 + 1271),

hí·ši?i·ruc *they're always red* (43 + S + 1071 + 1201.2 + 1271), kitó?ohi?i·ruc *they always turn green* (43 + 101 + S + 1071 + 1201.2 + 1271), té·kša?i·c *he always faints* (43 + S + 1041 + 1071 + 1271), té·kša?i·ruc *they always faint* (43 + S + 1041 + 1071 + 1201.2 + 1271), iré?-ekša·ci?i·c *he always keeps on kind of talking* (43 + S + 1041 + 1011 + 1071 + 1271), hí·šicaki?i·ruški *they're always bright red* (43 + S + S + 1071 + 1201.2 + 1283), ré·?i·ware·c *he went, they say* (43 + S + 1071 + 1261 + 1271).

± 40 + 21 or 40 + S + 1071 + 1272 or (1253 ± 1279): wa·pá·hi?i·š *the one who kept singing* (21 + S + 1071 + 1272), awaká?i·š *I who always look* (43 + 41 + S + 1071 + 1272), wa·rúhta·?i·šta *a crazy fellow* (21 + S + 1071 + 1253), wa·rúhta·?i·šta?a *is he a crazy fellow?* (21 + S + 1071 + 1253 + 1279).

21 or 40 ± S + S + (1021 + 40) or 1101 + 40) ± 1011 or 1091 + 1071 ± 1201 + 1271: kú·ciwa·?i·c (43 + S + 1021 + 41 + 1071 + 1271) ~ kú·cihkiwa·?i·c (43 + S + 1101 + 41 + 1071 + 1271) *I make him take it from him always*, wa·rúticakihe·?i·ruc *they always eat well* (21 + S + S + 1021 + 43 + 1071 + 1201.2 + 1271), aré?ehe·raci?i·c *he gets kind of mad every now and then* (43 + S + 1021 + 43 + 1011 + 1071 + 1271), wi·hiráphe·tha?i·ruc *they keep me awake all night* (41 + S + 1021 + 43 + 1091 + 1071 + 1201.2 + 1271).

40 ± 101 + S ± 1061 or 1011 + 1071 + 1101 + 40 + 1271: wihkicákiria?i·hkiwa·c *I always make myself fussy* (41 + 101 + S + 1061 + 1071 + 1101 + 41 + 1271), kárawi?i·hke·c *he always makes him remember* (43 + S + 1071 + 1101 + 43 + 1271), kárawa·ci?i·hke·c *he always makes him remember a little bit* (43 + S + 1011 + 1071 + 1101 + 43 + 1271).

40 + S ± 1011 + 1071 + 1101 + 40 + 1071 + 1271: kárawa·ci?i·hke·?i·c *he always makes him remember a little bit* (43 + S + 1011 + 1071 + 1101 + 43 + 1071 + 1271), kárawi?i·hke·?i·c *he always makes him remember* (43 + S + 1071 + 1101 + 43 + 1071 + 1271). (Either of the two preceding sequence types may occur alternatively to this—kárawi?i·hke·?i·c ~ kárawi?i·hke·c ~ kárawihke·?i·c *he always makes him remember.*)

± 31 + 40 + S + 1278 + S + 1071 + 1251 or (1252 + 1278): atá?akhu·?i·to·k *if he coming up over always comes* (43 + S + 1278 + S + 1071 + 1252 + 1278), aru?-atá?akhu·?i·ru *at the place where he comes up, coming here* (31 + 40 + S + 1278 + S + 1071 + 1251).

Suffix 1072, -ahi-, occurs in sequences of the following types.

40 + 40 ± 140 + S + 1072 ± 1201 or 1261 + 1271 or 1278: hí·ahic *he drank it all at once* (43 + 43 + S + 1072 + 1271), awakáha?ac *we caught sight of him* (43 + 41 + S + 1072 + 1201 + 1272) (cp·awaká?ac *we look*), ikáhiware·c *he caught sight of him they say* (43 + 43 + S + 1072 + 1261 + 1271), wa·pahcákahic *I slice it* (43 + 41 + 144 + S + 1072 + 1271) (cp·wa·pahcakic *I cut it*), kíkšahak *dressing him all at once* (43 + 43 + S + 1072 + 1278).

40 + S + 1101 + 40 + 1072 ± (40 + 1191) ± 1201 + 1271 or 1279): ó·cihkiwa·hic *I took a little rest* (43 + S + 1101 + 41 + 1072 + 1278), ó·cihkira·hi?i *did you take a little rest* (43 + S + 1101 + 42 + 1072 + 1279), ó·cihkiwa·hiwiha?ac *we're going to take a little rest* (43 + S + 1101 + 41 + 1072 + 41 + 1191 + 1201 + 1271).

Suffix Class 1090–1150

Suffix 1091, -tha negative, occurs word finally preceded by a stem, or word medially preceded by a stem or suffixes 1021 plus 40 and followed by suffixes of Classes 1010–1140, 1070–1130, 1170, 40 plus 1190, 1200, 1220, 1250, 1260 or 1270.

Sequences in which suffix 1091 occurs are of the following types.

± other prefixes + 40 ± S + S + 1091 ±

1091, 1012 or (40 + 1191) ± 1201 ± 1270: wašákitha *my thumb* (41 + S + 1091) (cp. wašáki *my hand*), atíthac *that's no house* (43 + S + 1091 + 1271), ó·tithak *it's not ripe, not being ripe* (43 + S + 1091 + 1278), rá·khitha?ac *they didn't get back* (43 + S + 1091 + 1201 + 1271), wahkárache·thathašha *though I didn't refuse* (41 + S + 1091 + 1091 + 1274), waríahithara·c *I almost stopped breathing* (41 + S + 1091 + 1012 + 1271), i·wa·?i·káthac *he can't see through it* (11 + 21 + 43 + S + 1091 + 1271), wihá·withawa *when I cannot sleep* (41 + S + 1091 + 1273), tó?ošetha?ac *they can't do anything about it* (43 + 43 + S + S + 1091 + 1201 + 1271), oruwitó?ošetha?ac *they can't do anything to me* (31 + 41 + 43 + S + S + 1091 + 1201 + 1271).

40 + S + 1091 + 1171 or (1251 ± 1278): hú·thaha *before he came* (43 + S + 1091 + 1171), ó·titharu *when they're not ripe* (43 + S + 1091 + 1251), hešétharuk *if it's not that* (43 + S + 1091 + 1251 + 1278).

± other prefixes + 40 + S + 1021 or 1101 + 40 + 1091 ± 1071 ± 1201 or 1261 + 1271: i·wa·?i·káhe·thac *by it he blinds him* (11 + 21 + 43 + S + 1021 + 43 + 1091 + 1271), i·wa·?i·káhe·thaware·c *by it he blinds him, they say* (11 + 21 + 43 + S + 1021 + 43 + 1091 + 1261 + 1271), arukó·wihe·tha?ac *they will never stop* (31 + 43 + S + 1021 + 43 + 1091 + 1201 + 1271), wi·hiráphe·tha?i·ruc *they keep me awake all night* (41 + S + 1021 + 43 + 1091 + 1071 + 1201 + 1271), ré·hke·thac *he doesn't let him go/want him to go* (43 + S + 1101 + 43 + 1091 + 1271).

S + 1091 ± 1282: awahútha *not as deep* (S + 1091), kuré?etha *don't carry it!* (S + 1091 + 1282).

30 ± S + S + 1091 + 1201 ± 1272: o·ruwa?etho?o *some who have none* (31 + S + S + 1091 + 1201), aku?ó·titha?aš *the ones that are not ripe* (32 + S + 1091 + 1201 + 1272).

S + 1091 + 1222: rúwa·cathari *just that one (person)* (S + 1091 + 1222).

Suffix Class 1100

Suffix 1101, -hke· (.1), hki (.2) causative, diminutive, imitative, occurs word finally preceded by a stem, or word medially preceded by a stem or suffixes of Class 1010–1140, 1020 plus 40, 1060, and 1070–1130 and followed by suffixes of Classes 40, 1010–1140, 1070–1130, 1090–1150, 40 plus 1190, 1200, 1250, and 1270.

The point two alternant, -hki, occurs when suffix 1101 is immediately followed by 41.9 or 42.9 (first or second person actor); the point one alternant, -hke·, occurs elsewhere.

Sequences in which suffix 1101 occurs are of the following types.

± other prefixes + 40 ± other prefixes + S ± (1021 + 40) ± 1011 + 1101 + 40 ± 1072 ± (40 + 1191) ± 1201 or 1261 + 1271, 1278 or 1279: kúxtihkiwa·c *I made him help him* (43 + S + 1101.2 + 41.9 + 1271), i·wikihí·šihke·c *I'm blushing* (11 + 41 + 101 + S + 1101.1 + 43.9 + 1271), arukiré·che·hke·c *he will make him use up his own* (31 + 43 + 101 + S + 1021 + 43 + 1101.1 + 43.9 + 1271), kahúkhe·hke·c *he made him fool him* (43 + S + 1021 + 43 + 1101.1 + 43.9 + 1271), ó·cihkira·?ahi?i *did you take a little rest?* (43 + S + 1101.2 + 42.9 + 1072 + 1279), ó·cihkiwa·?ahiwiha?ac *we're going to take a little rest* (43 + S + 1101.2 + 41.9 + 1072 + 41 + 1191 + 1201 + 1271), kicákihka·k *he becomes good/becoming good* (43 + 101 + S + 1101.1 + 43.9 + 1278), axpíhke·ware·c *he left it over, they say* (43 + S + 1101.1 + 43.9 + 1261 + 1271); for additional examples see those listed under suffix 1021 above.

40 ± 101 + S ± 1061 ± 1071 + 1101 + 40 ± 1071 ± 1201 + 1271: kú·cihkiwa·?i·c *I made him take it from him always* (43 + S + 1101.2 + 41.9 + 1071 + 1271), ré·hkiwa·?i·ruc *we let them go always* (43 +

S + 1101.2 + 41.9 + 1201 + 1271), wihkicákiria?i·hkiwa·c *I always make myself fussy* (41 + 101 + S + 1061 + 1071 + 1101.2 + 41.9 + 1271); for additional examples see those listed under suffix 1071 above.

40 + S + 1101 + 40 + 1091 + 1271: ré·hke·thac *he doesn't want him to go/he doesn't let him go* (43 + S + 1101.1 + 43.9 + 1091 + 1271).

40 + S + 1101 + 1281 or 1282: ikáhka *show them!* (43 + S + 1101 + 1282), ikáhka·ra *you (plural) show them!* (43 + S + 1101 + 1281).

40 + 101 + S + 1101 + 40 + 1251 ± 1278: kicó·tahke·ru *when they start fading* (43 + 101 + S + 1101.1 + 43.9 + 1251), kiré·šahke·ruk *when it is no more (becomes nothing)* (43 + 101 + S + 1101.1 + 43.9 + 1251 + 1278).

± prefixes + S + 1101 ± 1201 ± 1271 or (1251 + 1274): wíahke *small woman* (S + 1101.1), wáhtihke *toy boat or car* (S + 1101.1), wištáhke *my glasses* (41 + S + 1101.1) (cp. wištá *my eye*), a·tíhka·?a *their small houses* (43 + S + 1101.1 + 1201), watawa·?akupóhoruwihke *my apple* (41 + 51 + 21 + 32 + S + 1101.1), wa·?akucí·rihka·?aš *oranges* (21 + 32 + S + 1101.1 + 1201 + 1272), wa·rákahke·c *it's my doll* (43 + 41 + S + 1101.1 + 1271) (cp. wa·ráka *my young*), i·?axúahke·rušha *even curtains* (11 + S + 1101.1 + 1251 + 1274) (cp. axúac *it covers it*).

40 + S + 1101 + 1011 + 1201 + 1271: atíhke·raca?ac *these small houses are no good* (43 + S + 1101.1 + 1201 + 1271).

Suffix Class 1120

Suffix 1121, -hti desiderative, occurs word finally preceded by a stem and word medially preceded by a stem and followed by suffixes of classes 1010–1140, 1200 and 1270.

Suffix 1121 occurs in sequences of the following types.

± 21 + 40 ± 31 + S + 1121 ± 1011 ± 1201 ± 1271: wihá·wihtic *I'm sleepy, I want to sleep* (41 + S + 1121 + 1271), tiríahta? ac *they like to run* (43 + S + 1121 + 1201 + 1271), iré?ehta·cic *he kind of likes to talk* (43 + S + 1121 + 1011 + 1271), wa·wú·tihtic *I like to eat* (21 + 41 + S + 1121 + 1271), aruhiráwihti *his sleepy state* (41 + 31 + S + 1121), aruhiráwihto?o *their sleepy state* (41 + 31 + S + 1121 + 1201).

S + 1121 + 1282: hírihtika *try to do it!* (S + 1121 + 1282.1).

Suffix Class 1170

Suffix 1171, -ha in reference to time, place, occurs word finally preceded by a stem or suffixes of Classes 1050 or 1090–1150, and word medially preceded by a stem or suffix of Class 1090–1150 and followed by suffixes of Classes 1200, 1230, 1250 and 1270.

Suffix 1171 occurs in sequences of the following types.

40 ± S + S ± 1091 + 1171 ± 1200 ± 1278: wahkúha *while she was there* (43 + S + 1171), ácaha *while he was close/closely* (43 + S + 1171), ácahak *it being close* (43 + S + 1171 + 1278), ó·theriho?o *when they're almost ripe* (43 + S + S + 1171 + 1201), hú·thaha *before he came* (43 + S + 1091 + 1171), wa·ríšithahapa *before they dance* (43 + S + 1091 + 1171 + 1202).

S + 1171 + 1231 or 1251: é·raha·ru *long ago/past* (S + 1171 + 1251), é·raha·kua *in the past* (S + 1171 + 1231).

S ± 1051 + 1171 ± S: kixhúha *five times* (S + 1171), kixhúhapiraka *fifty* (S + 1171 + S), túawiha *how many times* (S + 1051 + 1171).

Suffix Class 1190

Suffix 1191, -hi (.1), -zero (.2), future, probability, ability, occurs word medially immediately preceded by suffixes of Class 40 (person actor) and followed by suffixes of Classes 1200 and 1270.

The point two alternant, -zero, occurs when suffix 1191 is immediately preceded by 41.10 or 42.10 (first or second person) and followed by suffixes other than 1201

(plural); the point one alternant, -hi, occurs elsewhere.

The actor is marked twice in forms containing suffix 1191, once immediately preceding 1191 (by 40.10) and once by prefix or, in the case of forms containing 1021 or 1101 (causatives), by suffix (40.9).

Suffix 1191 occurs in sequences of the following type.

± 40 + 40 + S ± (1021 + 40) ± 1011 ± (1101 + 40) ± 1072 or 1091 + 40 + 1191 ± 1201 + 1271: ácahic: *it might be near* (43 + S + 43 + 1191.1 + 1271), raráhuric *you must come* (42 + S + 42 + 1191.2 + 1271), wahé·wic *I might—/I can make it* (43 + 41 + S + 41 + 1191.2 + 1271), wahpiwiha?ac *we'll take a bath* (43 + 41 + S + 41 + 1191.1 + 1201 + 1271), hára·tharic *you can divorce him* (43 + S 1021 + 42 + 1091 + 42 + 1191.2 + 1271), cixíhkiwa·wic *I'll make him jump* (43 + S + 1101 + 41 + 41 + 1191.2 + 1271), ó·cihkiwa·?ahiwiha?ac *we're going to take a little rest* (43 + S + 1101 + 41 + 1072 + 41 + 1191.1 + 1201 + 1271), kárawihe·racihke·hic *he'll make him remember it a little bit* (43 + S + 1021 + 43 1011 + 1101 + 43 + 43 + 1191.1 + 1271).

Suffix Class 1200

Suffix Class 1200 (1201.-a?a .1, -o?o .2, -ru .3 plural; 1202. -apa group plural) occurs word finally preceded by a stem or suffixes of Classes 1000–1210, 1090–1150, 1120 or 1170, and word medially preceded by a stem or suffixes of Classes 40, 1010–1140, 1020, 1040, 1060, 1070–1130, 1090–1150, 1100, 1120 and 1190 and followed by suffixes of Classes 1000–1210, 1220, 1230, 1250, and 1270.

The point three alternant of 1201 (-ru) occurs when 1201 is immediately preceded by 1071 (durative, iterative); the point two alternant (-o?o) occurs when 1201 is word final or followed by suffix 1279 (glottal stop -repeat final mora); the point one alternant (-a?a) occurs elsewhere.

Sequences in which suffix 1201 occurs are of the type ± prefixes ± S + S ± other suffixes + 1201 ± 1001 or 1251 ± 1270, as; awó?o *lands* (S + 1201.2), í·xpo?o *his wings, their wings* (43 + S + 1201.2), wú·aš *fish* (S + 1201.1 + 1272), cácka?ac *they're noisy* (43 + S + 1201.1 + 1271), táha·?ac *they killed him* (43 + S + 1021 + 43 + 1201.1 + 1271), kí?iria?a?ac *they ride horseback* (43 + S + 1061 + 1201.1 + 1271), wa·?áhtu·?ahe *these skulls* (21 + S + 1201.1 + 1221), wa·há·tha?aru *when we cut it, when I cut them* (43 + 41 + 145 + S + 1201.1 + 1251), rawáha?aruk *if you (plural) want it* (43 + 42 + S + 1201.1 + 1251 + 1278), wi·cáha?ahta *by my canes* (41 + S + 1201.1 + 1001), aruhiráwihto?o *their sleepy state* (43 + 31 + S + 1121 + 1201.2), atíhko·?o *small houses* (S + 1101 + 1201.1), watawa·?akucí·rihka·?aš *my oranges* (41 + 51 + 21 + 32 + S + 1101 + 1201.1 + 1272), kitó?ohicaki?i·ruc *they always turn a good green* (43 + 101 + S + S+ 1071 + 1201.3 + 1271), híšicaki?i·ruški *they are always bright red* (43 + S + S + 1071 + 1201.3 + 1283), ré·hkiwa·?·ruc *we let them go always* (43 + S + 1101 + 41 + 1071 + 1201.3 + 1271).

Suffix 1202, -apa group plural, occurs in sequences of the following types.

± 40 + 40 + S ± 1012 + 1202 ± 1231 + 1270: á?apak *they have it, having it* (43 + 43 + S + 1202 + 1278), íapakuac *they're over there* (43 + S + 1202 + 1231 + 1271), tára·pac *they're dying* (43 + S + 1012 + 1202 + 1271).

40 + S + 1091 + 1171 + 1202: wa·ríši-thahapa *before they dance* (43 + S + 1091 + 1171 + 1202).

Suffix Class 1220

Suffix Class 1200 (1221. -he *that one here*, 1222. -ri objectifier, demonstrative, 1223. -ka *that one*, time-place localizer) occurs word finally preceded by stems or suffixes of Classes 1000–1210, 1010–1140, 1100, or 1200, and word medially preceded by stems or suffixes of Class 1200 and followed by suffixes of Classes 1230 and 1270.

Sequences in which suffix 1221 occurs are of the type ± 21 ± 11 + S ± 1201 + 1221 ± 1231, as: wacé·he *that man* (S + 1221), wá·rahe *this winter* (S + 1221), wa·ʔáhtu·ʔahe *these skulls* (21 + S + 1201 + 1221), wa·ʔi·ʔaxíhekua *in that corral* (21 + 11 + S + 121 + 1231).

Suffix 1222 occurs in sequences of the following types.

± 31 or (40 + 50) + S + 1222 ± 1270: ihá·ri *the other* (S + 1222) (cp.ihá·c *it's different*), ciríari *a cold* (S + 1222) (in ciríari wiʔákšiac *I caught a* cold), watxúpa·ri *my medicine* (41 + 51 + S + 1222) (cp. xúpa·c *it's holy*), orukíʔiri *the rider* (31 + S + 1222), šéʔeriwa (S + 1222 + 1273) or šéʔeric (S + 1222 + 1271) *that's the one.*

S + 1001 or 1201 + 1222: ihá·htari *in, toward the other* (S + 1001 + 1222), šíʔari *those* (S + 1201 + 1222).

Suffix 1223 occurs in sequences of the following types.

± 31 ± S + S ± 1201 + 1223 ± 1231: awahúka *inside, in the inside* (S + 1123), wapíhirika *nowadays* (S + S + 1223), aruwatúka *where there is* (31 + S + 1223) (in wirí + aruwatúka *where there is water*), síʔakakua *at that time* (S + 1201 + 1223 + 1231).

40 + S + 1001 + 1101 + 1201 + 1223: iwaráʔacihkaʔaka *we just talk about it* (43 + 41 + S + 1001 + 1101 + 1201 + 1223).

Suffix Class 1230

Suffix Class 1230 (1231. -ku .1, -hkua .2 *in, into, that one*; 1232. -kuha· .1, -hkuha· .2 *from*) occurs word finally preceded by stems or suffixes of Classes 1000–1210, 1170, 1200 or 1220, and word medially preceded by stems or suffixes of Class 1200 and followed by suffixes of Classes 1250 and 1270.

The point two allomorphs (1231.2 -hkua and 1232.2 -hkuha·) occur when suffixes 1231 and 1232 are preceded by long vowels or identical vowels separated by glottal stop; the point one allomorphs (1231.1 -kua and 1232.1 -kuha·) occur elsewhere. (For the alternation of morpheme final /a·/ with /a/ when followed by juncture see **1.3.**)

Suffix 1231 occurs in sequences of the following types.

± (21 ± 11) + S ± 1201 ± 1220 + 1231: atíkua *in the house* (S + 1231.1), wa·ʔákašikua *in the paper* (21 + S + 1231.1), wirá·ʔahkua *into the fire* (S + 1201 + 1231.2), šíʔakakua *at that time* (S + 1201 + 1223 + 1231.1), wa·ʔi·ʔaxíhekua *in that corral* (21 + 11 + S + 1221 + 1231.1).

S + 1001 or 1171 + 1231: réhtakua *along the edge* (S + 1001 + 1231.1), é·rahakua *long ago* (S + 1171 + 1231.1).

40 + S ± 1202 + 1231 ± 1251 + 1270: íapakuac *they are there* (43 + S + 1202 + 1231.1 + 1271), tá·pakuaruk *wherever she is (going)* (43 + S + 1231.1 + 1251 + 1278).

Sequences in which suffix 1232 occurs are of the following type: ± 40 + S + 1232 ± 1278: šéʔehkuha *from that, for that reason* (S + 1232.2), í·hkuha *from her mouth* (43 + S + 1232.2), páhcitakuha·k *from the west, he being from the west* (43 + S + 1232.1 + 1278).

Suffix Class 1240

Suffix 1241, -ši in reference to time, occurs word finally preceded by a stem or word medially preceded by a stem or a suffix of Class 40 (preceded by 1021) and followed by suffixes of Classes 1250 or 1270.

Sequences in which suffix 1241 occurs are of the following types.

31 + S + 1241: arukixhúši *at five* (31 + S + 1241).

S + 1241 + 1251: ó·kciaširu *last night* (S + 1241 + 1251), hú·riširu *yesterday* (S + 1241 + 1251).

40 + S + 1241 + 1276: arí·tišihiri *because he's hungry* (S + 1251 + 1276), tá·pašihiri *that's why* (S + 1251 + 1276).

40 + S + S + 1021 + 40 + 1241 + 1273: ixúare·he·šiwa *because it made him sick* (43 + S + S + 1021 + 43 + 1241 + 1273).

Suffix Class 1250

Suffix Class 1250 (1251. -ru inclusive (time or place) *if*; 1252. -to·- indefinite, *or*; 1253. -šta· ~ -št habitual agent, past definite) occurs word finally preceded by stems or suffixes of Classes 40, 1020, 1070–1130, 1090–1150, 1100, 1170, 1200, 1220 and 1240, or word medially preceded by stems or suffixes of Classes 40, 1020, 1070–1130, 1090–1150, 1100, 1200, and 1230 and followed by suffixes of Class 1270.

The variation between the two allomorphs of 1253 (-šta·, -št) is free.

Suffix 1251, -ru, occurs in sequences of the following types.

± 40 + 40 ± other prefixes + S ± (1091 ± 1223) or 1201 or 1223 + 1251 ± 1270: ó·tiru *when they're ripe* (40 + S + 1251), héšetharuk *if it's not that* (40 + S + 1091 + 1251 + 1278), rawahá?aruk *if you (plural) want* (43 + 42 + S + 1201 + 1251 + 1278), tá·pakuaruk *wherever she is (going)* (43 + S + 1231 + 1251 + 1278), wakúrušha *even if I give him* (43 + 41 + S + 1251 + 1274), ri?a·ráciruc *he's behind you* (42 + 43 + S + 1251 + 1271), kitúaru *when it turns green* (43 + 101 + S + 1251), ó·tithakaru *when they're not ripe* (43 + S + 1091 + 1223 + 1251), hí·ruk *if he came* (43 + S + 1251 + 1278).

101 + S + 1101 + 40 + 1251 ± 1278: kicó·tahke·ru *when it starts fading* (101 + S + 1101 + 43 + 1251), kiré·šahke·ruk *when it is no more* (101 + S + 1101 + 43 + 1251 + 1278).

± (40 + 50) ± other prefixes + S ± 1021, 1101, 1171 or 1241 + 1251 ± 1274: a·táru *in the morning, tomorrow* (S + 1251), wi·té·rušha *even buffalo* (S + 1251 + 1274), é·raharu *long ago* (S + 1171 + 1251), ó·kciaširu *last night* (S + 1241 + 1251), i·?axúahke·rušha *even curtains* (11 + S + 1101 + 1251 + 1274), ita?aru?apaxíhe·ru *his place for resting* (43 + 51 + 31 + S + 1021 + 1251).

31 + 40 + S + 1278 + S + 1071 + 1251: aru?atá?akhu·?i·ru *where he always comes up, coming here* (31 + 43 + S + 1278 + S + 1071 + 1251).

Suffix 1251, -to·- occurs in sequences of the following types.

± 40 ± 40 + S + 1252 + 1278: kixhúto·k *about five* (S + 1252 + 1278), rihá·wito·k *are you going to sleep, or* — (42 + S + 1252 + 1278), šé?eto·k—šé?eto·k *whether it's that one—or that one* (43 + S + 1252 + 1278—43 + S + 1252 + 1278), kiwaxhúto·k *perhaps he asked him* (43 + 43 + S + 1252 + 1278).

40 + S + 1278 + S + 1071 + 1251 + 1278: atá?akhu·?i·to·k *he comes up, coming here always, I wonder* (43 + S + 1278 + S + 1071 + 1251 + 1278).

Suffix 1253, -šta· ~ -št, occurs in sequences of the following type.

± 40 ± other prefixes + S ± 1021 or 1071 or (1101 + 40) ± 1201 + 1253 ± 1279: té·šta· *he had died* (43 + S + 1253), wapašá·ka?ašt *we already cut it* (43 + 41 + 144 + S + 1201 + 1253), wa·rúhta·?i·šta·?a *is he a crazy fellow?* (21 + S + 1071 + 1253 + 1279), akutáhe·šta³ *the one who was killed* (32 + S + 1021 + 1253), akutáhe·šta·?a *which one is it who was killed?* (32 + S + 1021 + 1253 + 1279), ré·hkiwa·št *I did let him go* (43 + S + 1101 + 41 + 1253), ré·hkiwa·?ašt *we already let him go* (43 + S + 1101 + 41 + 1201 + 1253).

Suffix Class 1260

Suffix 1261, -ware·- quotative, occurs word medially preceded by stems or suffixes of Classes 40, 1020, 1070–1130, 1090–1150, 1100 or 1200 and followed by suffix 1271.

Sequences in which suffix 1261 occurs are of the following type.

± other prefixes + 40 ± 101 ± 140 + S ± (1021 or 1101 + 40) ± 1071 or 1091 ± 1201 + 1261 + 1271: wi?í·watuware·c *I was born* (41 + S + 1261 + 1271), hašhirá·?aware·c *they are grizzlies, they say* (43 + S + 1201 + 1261 + 1271), kiruškíware·c *he pulled out his own, they say* (43 + 101 + 141 + S + 1261 + 1271), ré·?i·ware·c *he*

always went, they say (43 + S + 1071 + 1261 + 1271), ré·ʔi·ruware·c *they always went, they say* (43 + S + 1071 + 1201 + 1261 + 1271), i·wa·i·káthaware·c *he can't see through it, they say* (11 + 21 + 43 + 43 + S + 1091 + 1261 + 1271), i·wa·ʔi·káhe·thaware·c *by it he blinds him, they say* (11 + 21 + 43 + S + 1021 + 43 + 1091 + 1261 + 1271), axpíhke·ware·c *he left it over, they say* (43 + S + 1101 + 43 + 1261 + 1271).

Suffix Class 1270–80

The members of suffix class 1270–80 are: 1271.-c predicative, utterance final; 1272.-š substantive, *named*; 1273.-wa *as, when, at*; 1274.-ša concessive; 1275.-še distributive; 1276.-hiri *because*; 1277.-rahe reportative; 1278.-ak (.1), -k (.2) predicative, non-utterance final; 1279.-ʔ repeat-final-mora interrogative; 1281.-ara imperative plural; 1282.-zero ~ -h ~ -ka (.1), -a ~ -ah ~ -aka (.2) imperative singular; and 1283.-ški emphatic.

Suffix class 1270–80 occurs word finally preceded by stems and suffixes of all other classes except 1050 (quantifier). (Suffix 1051 occurs so infrequently in our corpus as to make its potential associations uncertain.)

Suffix 1271.-c, utterance final predicative, occurs in sequences of the type: ± other prefixes + 40 or 101 ± other prefixes + S ± other suffixes + 1271: wihá·wic *I sleep* (41 + S + 1271), riʔakuwa·híric *you're a worker* (42 + 32 + 21 + S + 1271), kiré·che·c *he uses up his own* (101 + S + 1021 + 43 + 1271), ísapakuac *they are there* (43 + S + 1202 + 1231 + 1271); for additional examples see those listed under all affix classes except 1050, 1170, and 1240.

Suffix 1272, substantive, *named* occurs in sequences of the following type: ± prefixes + S ± other suffixes + 1272: arí·š *the road* (S + 1272), cáckiš *Noizy* (personal name—S + 1272), wú·aš *fish* (S + 1201 + 1272), wa·pá·hiʔi·š *the one who kept singing* (21 + S + 1071 + 1272); for additional examples see those listed under Classes 30, 1010, 1090 and 1200.

Suffix 1273, -wa *as, when, at*, occurs in sequences of the type: ± prefixes + S ± other suffixes + 1273: ó·kciawa *at night* (S + 1273), té·wa *when she died* (43 + S + 1273), ká·ʔawa *as they have been there* (43 + S + 1201 + 1273); for additional examples, see those listed under Classes 30, 1020, 1090, 1200, 1220 and 1240.

Suffix 1274, -ša concessive, occurs in sequences of the type: ± prefixes + S ± other suffixes + 1274: atíša *a house, but—* (S + 1274), witíaša *although I'm late* (41 + S + 1274), hu·párusha *even shoes* (S + 1251 + 1274); for additional examples, see those listed under Classes 1010, 1090, 1100 and 1250.

Suffix 1275, -še distributive occurs in sequence of the type: ± prefixes + S ± 1201 + 1275: atíše *camp* (S + 1275, cp. atí *house*), a·tá·ʔaše *their several houses* (41 + S + 1201 + 1275), aru·ʔú·ʔuše *places where wounded* (31 + S + 1275).

Suffix 1276.-hiri *because* occurs in sequences of the type: 40 + S + 1241 + 1276: tápa·šihiri *because it's soft* (41 + S + 1241 + 1276), arí·tišihiri *because he's hungry* (41 + S + 1241 + 1276).

Suffix 1277, -rahe reportative, occurs in sequences of the type: 40 + S ± 1201 + 1277: hiráwirahe *he slept, so reported* (43 + S + 1277), há·ʔarahe *they said so, so reported* (43 + S + 1201 + 1277).

Suffix 1278, -ak (.1), -k (.2) non-utterance final predicative, occurs in sequence of the type: ± other prefixes + 40 ± other prefixes + S ± other suffixes + 1278: ó·tak *it's ripe/it being ripe* (43 + S + 1278.1), kicákihka·k *he becomes good/becoming good* (43 + 101 + S + 1101 + 43 + 1278.1), rawahá·ʔaruk *if you (plural) want* (43 + 42 + S + 1201 + 1251 + 1278.2), rihá·wito·k *are you going to sleep, or—* (42 + S + 1252 + 1278.2). The point two alternant, -k, occurs when 1278 is preceded by a morpheme ending in /o/ or /u/; the point one alternant, -ak, occurs elsewhere.

Suffix 1279, -ʔ repeat-final-mora interrogative, occurs in sequences of the type: ± prefixes + S ± other suffixes + 1279: tá·paʔa *what?* (S + 1279), ricawé·ʔe *are you hot?* (42 + S + 1279), ó·cihkira·ʔahiʔi *did you take a little rest?* (43 + S + 1101 + 42 + 1072 + 1279).

Suffixes 1281, -ara imperative plural, and 1282, -zero ~ -h ~ -ka (.1), -a ~ -ah ~ -aka (.2) imperative singular, occur in sequences of the type: 40, 21 or 101 + S ± 1021, 1101, or 1121 ± 1091 + 1281 or 1282: tá *die!* (43 + S + 1282.2), tá·ra *you (plural) die!* (43 + S + 1281), ripáhtah *roll over!* (42 + S + 1282.2), ripáhta·ra *you (plural) roll over!* (42 + S + 1281), wa·pá·ha *sing something!* (21 + S + 1282.2), wa·pá·ha·ra *you (plural) sing something!* (21 + S + 1281), árakitha *set it on fire!* (43 + S + 1021 + 1282.2), wa·pháhka *motion!* (21 + S + 1101 + 1282.2), ké?etha *don't dig!* (43 + S + 1091 + 1282.2), ko·wíhe·tha *don't quit!* (43 + S + 1021 + 1091 + 1282.2), tá·t *cross!* (43 + S + 1282.1), hú ~ hú·h *come!* (43 + S + 1282.1), kíruph ~ kíruwika *count it!* (43 + S + 1282.1), hací·tha ~ hací·thah ~ hací·tha·ka *roast it!* (43 + S + 1282.2), hírihtika *try to do it!* (43 + S + 1121 + 1282.1).

The variation between the alternants of 1282 is in part morphemically determined and in part free, that among the sub-alternants (separated by the alternation sign) is free—i.e. -zero, -h and -ka of point one vary freely, as do -a, -ah and -aka of point two. Alternant point one occurs when suffix 1282 is immediately preceded by stems ending in weak vowels (see **1.1.1**). Alternant point also occurs after certain other stems for which no generalization in terms of phonemic shape is possible. (Such stems which appear in our sample are listed in **3.**). Alternant point two occurs when suffix 1282 is preceded by other stems (listed in **3.**).

Suffix 1283, -ški emphatic, occurs in sequences of the type: 40 ± S + S ± 1071 ± 1201 + 1283: ré·šaški *there's nothing at all* (43 + S + 1283), tóʔohiʔi·ški *it is always blue* (43 + S + 1071 + 1283), hí·šicakiʔí·ruški *they are always bright red* (43 + S + S + 1071 + 1201 + 1283).

2.3.1. Particular alternants of each of the person marking morphemes (the members of Class 40-90-110-1030-1110-1180) may be combined into sub-sets on the basis of their occurrence with particular themes or other affixes, or their position in respect to themes or other affixes. Each of these sub-sets (hereinafter to be called sets) contains three members—one alternate of each of the three morphemes. The alternants are numbered on the basis of their membership in these sets.

Thus, the sets are: Set Point One (.1): wi- ~ wih- (41.1), ri- ~ rih- (42.1), i- ~ ih- (43.1).

Set Point Two (.2): wi·- (41.2), ri·- (42.2), zero- (43.2).

Set Point Three (.3): wa- (41.3), ri- (42.3), i- (43.3).

Set Point Four (.4): wa- (41.4), ri- (42.4), zero- (43.4).

Set Point Five (.5): wa- (41.5), ra- (42.5), a- (43.5).

Set Point Six (.6): wa- ~ wah- (41.6), ra- ~ rah- (42.6), zero- (43.6).

Set Point Seven (.7): w- (41.7), r- (42.7), zero- (43.7).

Set Point Eight (.8): wi- (41.8), ri- (42.8), zero- (43.8).

Set Point Nine (.9): -wa· (41.9), -ra· (42.9), -zero (43.9).

Set Point Ten (.10): -wi (41.10), -ri (42.10), -zero (43.10).

The overlapping in set membership of particular alternants, as shown by the multiple numbering of those alternants, in **2.1**, above, may be summarized as follows. For 61: the alternant wi- occurs as a member of Sets Point One and Point Eight; the alternant wa- occurs as a member of Sets Point Three, Point Four, Point Five and Point Six; the alternants wah-, wi·-, w-, -wa·, and -wi occur as members of only one set each, namely, Sets Point Six, Point Two,

Point Seven, Point Nine, and Point Ten respectively. For 62: the alternant ri- occurs as a member of Sets Point One, Point Three, Point Four, Point Eight; the alternant ra- occurs as a member of Sets Point Five and Point Six; the alternants rah-, ri·-, r-, ra· and -ri occur as members of only one set each, namely, Sets Point Six, Point Two, Point Seven, Point Nine and Point Ten, respectively. For 63: the alternant i- occurs as a member of Sets Point One and Point Three; the alternant zero prefix occurs as a member of Sets Point Two, Point Four, Point Six, Point Seven and Point Eight; the alternant zero suffix occurs as a member of Sets Point Nine and Point Ten; the alternant a- occurs only as a member of Set Point Five.

2.3.2. The particular theme or class of themes with which each set occurs may be termed its invariant. The grammatical meaning or function of the set plus its invariant—or rather, plus the class meaning of the invariant—may be termed the constant for that set.

Certain sets, as shown below, may be combined into sub-systems of isosemantic sets, having common constants. Each of these larger sets is called a paradigm and assigned a letter. So also, those single sets having unique constants are called paradigms and assigned letters.

2.3.2.1. Paradigm A. This paradigm consists of Sets Point One, Point Two, Point Three, Point Four, Point Five, Point Six and Point Seven, for which in sequence with noun themes, plus or minus other affixes which do not include ta- alienable (51), the constant is inalienable possessor.

The occurrence of a particular set with a particular theme or class of themes is stateable only in terms of lists. This means that it is impossible to obtain exhaustive generalizations in terms of the phonemic or morphemic constituents of the themes; however, one restricted generalization is possible: themes with initial /u/ or /u·/ occur only with Set Seven (w-, r-, zero-).

Examples of themes occurring with each set of paradigm A are listed below.

Set Point One: wiwáki *my chest*, riwáki *your chest*, iwáki *his chest*. Set Point Two: wi·cí·ta *my tail*, ri·cí·ta *your tail*, cí·ta *his tail*. Set Point Three: wací *my foot*, rici *your foot*, ici *his foot*. Set Point Four: wašáki *my hand*, rašáki *your hand*, šáki *his hand*. Set Point Five: wapíša *my liver*, rapíša *your liver*, apíša *his liver*. Set Point Six: wakirá?aš *my husband*, rakirá?aš *your husband*, kirá?aš *her husband*. Set Point Seven: wé?e *my food*, ré?e *your food*, é?e *his food*.

2.3.2.2. Paradigm B. This paradigm contains only Set Point Three, for which in sequence with ta- alienable (51) and noun themes, the constant is alienable possessor. As: wata?awá *my land*, rita?awá *your land*, ita?awá *his land*.

2.3.2.3. Paradigm C consists of Sets Point One, Point Two, and Point Eight, for which in sequence with verb themes the constant is stative actor.

Examples of themes occurring with each set of paradigm C are listed below.

Set Point One: wirú·hic *I stand up*, rirú·hic *you stand up*, irú·hic *he stands up*. Set Point Two: wi·xakác *I move*, ri·xakác *you move*, xakác *it moves*. Set Point Eight: wihapá·c *I'm cold*, rihapá·c *you're cold*, hapá·c *he's cold*.

2.3.2.4. Paradigm D consists of Sets Point Three, Point Five, Point Six and Point Seven, for which in sequence with verb themes the constant is active actor.

Examples of themes occurring with each set of paradigm D are listed below.

Set Point Three: waré?ec *I talk*, riré?ec *you talk*, iré?ec *he talks*. (This is the only theme in the sample which occurs with Set Point Three as active actor.) Set Point Five: wa?áciwic *I follow tracks*, ra?áciwic *you follow tracks*, a?áciwic *he follows tracks*. Set Point Six: wahkárathic *I step on it*, rahkárathic *you step on it*, kárathic *he steps on it*; watiríac *I run*, ratiríac *you run*, tiríac *he runs*. Set Point Seven: wú?uc *I wound him*, rú?uc *you wound him*, ú?uc *he wounds him*.

2.3.2.5. Paradigm E consists of Sets Point Two and Point Eight and the prefix 41 (first person) member of Set Point Seven, for which in sequence with themes plus Paradigm D (active actor) or suffixes 1281 or 1282 (imperative) the constant is goal.

Examples of themes occurring with each set of paradigm E are listed below.

Set Point Two: wi·kúxtic *he helps me* (41.2 + 43.6 + T + 1271), ri·kúxtic *he helps you* (42.2 + 43.6 + T + 1271), kúxtic *he helps him* (43.2 + 43.6 + T + 1271). Set Point Eight: wikúc *he gives me* (41.8 + 43.6 + T + 1271), rikúc *he gives you* (42.8 + 43.6 + T + 1271), kú?uc *he gives him* (43.8 + 43.6 + T + 1271). (For the alternation of ku- and ku?u- see **1.4.**) Set Point Seven: wúk *give me!* (41.7 + T + 1282). (This is the only example in the total corpus obtained of a member of Set Point Seven as goal, and here it occurs with an alternant of the theme which occurs only in this position—i.e. when preceded by 41.7 as goal—see **1.4.**)

2.3.2.6. Paradigm F contains only Set Point Nine, for which in sequence with any theme plus suffixes 1021 or 1101 (causative) the constant is causative actor. As: cá·khiwa·c *I make him wail* (43.2 + T + 1021.2 + 41.9 + 1271), cá·khira·c *you make him wail* (43.2 + T + 1021.2 + 42.9 + 1271), cá·khe·c *he makes him wail* (43.2 + T + 1021.1 + 43.9 + 1271); kúxtihkiwa·c *I made him help him* (43.2 + T + 1101.2 + 41.9 + 1271), kúxtihkira·c *you made him help him* (43.2 + T + 1101.2 + 42.9 + 1271), kúxtihke·c *he made him help him* (43.1 + T + 1101.1 + 43.9 + 1271).

2.3.2.7. Paradigm G contains only Set Point Ten, for which in sequence with a theme plus suffix 1191 (future) the constant is future actor. As: cixíhkiwa·wic *I'll make him jump* (43.2 + T + 1101.2 + 41.9 + 41.10 + 1191.2 + 1271), cixíhkira·ric *you'll make him jump* (43.2 + T + 1101.2 + 42.9 + 42.10 + 1191.2 + 1271), cixíhke·hic *he'll make him jump* (43.2 + T + 1101.1 + 43.9 + 43.10 + 1191.1 + 1271).

2.3.3. Paradigms A, B, and C (inalienable and alienable possessor and stative actor) occur in decade position 40 or 100, before or after Decades 10–60, 20–70, and 30–80.

Paradigm D (active actor) occurs in decade position 100, preceded by Paradigm E (goal) in decade position 90, except in warikúc *I give you* (41.6 + 42.8 + T + 1271), in which Paradigm D precedes Paradigm E.

Paradigm E (goal) occurs in decade position 90 (except in warikúc, above), and always occurs in sequence with Paradigm D (active actor), Paradigm F (causative actor), Paradigm G (future actor), or suffix 1281 or 1282 (imperative).

Paradigm F (causative actor) occurs in decade positions 1030 and 1110, immediately preceded by suffix 1021 (when Paradigm F is in position 1030) or suffix 1101 (when Paradigm F is in position 1110).

Paradigm G (future actor) occurs in decade position 1180, immediately followed by suffix 1191 (future).

2.3.4. The overlapping in paradigm membership of particular sets may be summarized as follows. Set Point One appears in Paradigms A and C (inalienable possessor and stative actor). Set Point Two appears in Paradigms A, C and E (inalienable possessor, stative actor and goal). Set Point Three appears in Paradigms A, B and D (inalienable possessor, alienable possessor and active actor). Set Point Four appears only in Paradigm A (inalienable possessor). Sets Point Five and Point Six appear in Paradigms A and D (inalienable possessor and active actor). Set Point Seven appears in Paradigms A and D (inalienable possessor and active actor) and the first person member appears in Paradigm E (goal). Set Point Eight appears in Paradigms C and E (stative actor and goal). Sets Point Nine and Point Ten appear in one paradigm each, Paradigm F (causative actor) and Paradigm G (future actor), respectively.

NOTES AND REVIEWS

THREE KEYS TO LANGUAGE. By Robert M. Estrich and Hans Sperber. New York: Rinehart & Co., Inc., 1952.

FRANKLIN FEARING

Three Keys to Language is distinguished from many other discussions of the subject by the intent to examine the phenomena of language in the context of human behavior as a whole. The authors view the problems of linguistics in terms of an intricate interplay of psychological, social, and historical factors. These are the 'three keys', but since they cannot be sharply set apart from each other, they actually turn out to be one 'key', namely, psychology, or, more strictly, social psychology. Regardless, however, of the labels applied to these factors, it is the essence of the view that there are many forces, perhaps the major forces, outside of language itself which shape it, determine how it is used, and its course of change.

Proceeding from this general point of view the authors find themselves, of course, involved in the several fields not usually regarded as part of linguistic study. Utilizing the perspectives furnished by cultural history, political history, anthropology, and particularly psychology, they deal with a variety of linguistic phenomena. These include word taboos in primitive and civilized societies, speech communities, phonetic, word, and semantic change, personal and social factors in the determination of language styles, humor in language, and standards of 'good' usage.

This reviewer is not prepared to say to what extent such an approach is regarded as heretical by the linguistic specialists, but it is refreshing to find the study of language extended beyond the somewhat narrow confines of conventional linguistic methodology. The authors are quite specific on this point. Linguistics means 'the study of all the problems of language' and its range may not be limited to those 'that can be approached by the methods of natural science'. This last phrase may be taken to mean that merely because a problem is not amenable to particular methods (presumably those of conventional linguistic study) it is not a legitimate part of scientific linguistics.

This is excellent. Such an approach, however, demands a better than average familiarity with the methods and concepts of the various behavioral sciences. It must be admitted that any attempt to transcend the wholly artificial boundaries that mark off the traditional fields of knowledge is always beset with difficulties, not the least of which is the tendency of those whose fields are invaded to take potshots at the invader. But in the present instance, admirable as it is in intent, and, in a measure, in execution, the undertaking rests so heavily on concepts from a particular discipline, we are bound to scrutinize those concepts carefully, not to repel the invader, but to insure his ultimate success.

The discipline is, of course, psychology. 'Language', say the authors, 'is largely a matter of psychology'. Although committed to its use, they are not altogether easy with the term. They feel impelled to state that the reader has the right to insert a skeptical 'whatever that may mean' whenever the word is used. In order to escape the ambiguity which, they believe, attaches to the term and the 'variety of notions' which it embraces, they specify its meaning in the context of linguistics. It refers to two basic things. First, any individual use of language is the outcome of individual conditions and behavior, and second, any development of language in the group, although it begins as an individual act, must achieve acceptance in the speech community. This implies, in other words, that the whole psychological

structure of the individual—his needs, motives, beliefs, etc.—determine what and how he will use words, and that if these linguistic usages are accepted by the group they must somehow meet the needs of a speech community.

This is fine as far as it goes. But instead of utilizing current findings and theory in the field of cognition-perception, the authors fall back on the outmoded concepts of human behavior in which emotional and rational factors are seen as distinct and opposed forces. There is no evidence in the book that the writers were aware of the work of Gardner Murphy on the development of perception, Heinz Werner's extremely fruitful concepts of the nature of cognitive and perceptual development in children and primitives, Jean Piaget's extended studies of thought processes in relation to language in children, Wayne Dennis' work on conceptualization in children in literate and non-literate cultures, and the work of many social psychologists on the so-called 'social' factors in perception. This work bears directly on the problems in hand, and its consideration would have enormously reinforced the authors' announced intention of utilizing the resources of psychology in considering the problems of linguistics.

Language is largely a matter of psychology. The crucial problem is the nature of the relation of the cognitive-perceptual process to the acquisition and use of linguistic techniques. The relationship is intimate and subtle. The Whorfian thesis that the linguistic patterns not only determine the cognitive processes but are prior to them needs further substantiation, but his analyses of the relationships between the way the individual perceives and thinks about his world and his linguistic patterns constitutes a point of departure for all those concerned with the 'psychology of' language.

In considering these problems it is not necessary to invoke the emotions or feelings as special forces which may be present or absent in language usage, or to separate the intellectual (logical) from the emotional use of words. Murphy has pointed out (Psych. Rev., 1951, 58: 332–349) that it is difficult if not meaningless to separate the cognitive, affective and conative aspects of perception. Insofar as language is part of the cognitive-perceptual processes, this would seem to imply that language, that is, language as used by human beings is also inescapably 'emotional'.

Estrich and Sperber clearly recognize the importance of emotional factors especially in semantic change. But one suspects that they can't quite shake off a conviction that the presence of such factors is somehow undesirable. Perhaps their difficulty arises from an unconscious commitment to the doctrine that the observed change in meaning is a change from an original meaning which was the 'true' or 'real' meaning, and that subsequent extensions, arising in an 'emotional' context are distortions from this. One detects a note of regret when the authors state that 'a really foolproof language is beyond human possibilities' because one of the basic tasks of language is 'to give vent to the emotions'. It has always been difficult to divest the word 'emotion' of undesirable connotations, especially when it is opposed to 'unemotional' or 'rational'.

As Kenneth Burke points out language always or nearly always expresses attitudes; it indicates an orientation. This is necessarily 'emotional'. Attitudes, whatever else they may be, have an affective core. In the essay entitled 'Semantic and Poetic Meaning' in The Philosophy of Literary Form he discusses at length the problem of the two modes of language usage—the 'semantic' or neutral mode which would probably be Estrich and Sperber's 'foolproof language', and the 'poetic' or emotional mode.

'The semantic ideal,' he says, 'would eliminate the *attitudinal* ingredient from its vocabulary (seeking a vocabulary for events equally valid for use by friends,

enemies, and the indifferent)'. Such an ideal, however, is itself 'an attitude, hence never wholly attainable, since it could be complete only by the abolition of itself.' (p. 150) While granting its incisiveness and the rather fictitious accuracy it confers, the semantic mode has but a limited usefulness, but 'above all' he adds, 'it fosters, sometimes explicitly, sometimes by implication, the notion that one may comprehensively discuss human and social events in a nonmoral vocabulary and that perception itself is a nonmoral act'. (p. 164)

Whether the Burkean definitions are accepted or not, the problem of emotional factors in language usage does not seem to be adequately met by a psychological theory which assumes that emotional and nonemotional factors are separable.

Any discussion of the social and psychological aspects of language usage sooner or later must concern itself with the intricate and baffling problems of metaphor. Effectively utilizing their extensive historical data collected in connection with the compilation of the Dictionary of Political Words and Phrases the authors are able to demonstrate the important role which metaphor plays in political speech. Their discussion of the historical origin and use in American political life of such terms as *copperhead*, *Loco-foco*, *carpet-bagger*, and the like, reveals the enormous potency that metaphor possesses to mobilize and unify men in political action. Here again, however, it seems to this reviewer that they have failed to discuss the relation of metaphor to the cognitive-perceptual processes in man. They assign three social functions to metaphor: (1) to enable the speaker to introduce his personal interests into a speech situation not of his own choosing, (2) to enable a speaker to awaken interest in an audience, and (3) to establish an intellectual or emotional speech community by exploiting common interests. While this deliberate employment of metaphor undoubtedly occurs, we are bound to ask how it is that particular words or phrases acquire their unique emotional and cognitive significance.

It is the essence of metaphor that something is "seen" in terms of something else. THIS establishes a perspective on THAT. It does not seem sufficient to understand this relationship as merely an additive one, or one which is the result of a voluntary decision—by a speaker, for example. Rather it would seem to be based on a type of immediate, total experience to be understood in terms of perceptual-cognitive processes, not as the result of conscious analysis and decision to associate THIS with THAT. As William Empson points out in The Structure of Complex Words metaphor is a matter of insight involving, of course, emotional components as do all cognitive processes.

The approach of the gestalt psychologists —especially Heinz Werner, Kurt Koffka and others—is useful here. Their conceptions of physiognomic perception and Werner's conception of syncretic imagery appear to be basically related to metaphor. The exact nature of the relationship between metaphorical language and physiognomic perception needs to be worked out. The approach of Estrich and Sperber with its avowed intent of including psychological, social, and historical factors sets the stage for this. It suggests that metaphor like other phenomena of language cannot be understood in purely linguistic terms.

There are many signs that a rapproachment between the various social science disciplines and linguistics is in the making. To this psychology has an especially important contribution to make. This is especially true of the kind of thinking associated with gestalt psychology and contemporary social psychology, sometimes called 'field theoretical' psychology. It may be, too, that the so-called learning theorists with their methodological precision and elegance may make an important contribution, although to this writer this is not so obvious. In any event, language, culture, and the psychological structure of the indi-

vidual are seen as interrelated phenomena, and, if for no other reason, Three Keys to Language is an important book because it presents a persuasive statement of this thesis.

University of California, Los Angeles

Report of the Fourth Annual Round Table Meeting on Linguistics and Language Teaching. Edited by Archibald A. Hill. Georgetown University Press, 1953. Pp. 115.

William E. Bull

The major objective of the Georgetown round-table meetings is to restore something of the molecular structure of the humanities which that process of disintegration called specialization has split into so many elementalistic groups that the linguist, the language teacher, and the man of letters not only do not talk the same language but operate in an atmosphere of mutual misunderstanding and distrust. In broad generalizations the linguists do not get along with the language teachers, the language teachers are allergic not only to linguistic research but to the practical applications of the results to class-room procedures, and many professors of literature tend to consider both the linguist and the language teacher, at best, as necessary deviates who justify departmental budgets and, at worst, as some type of subversive whose aim is to make language the exclusive concern of language departments.

The response of the Institute of Languages and Linguistics to this lamentable situation has been a series of round-table meetings based on the bold and constructive assumption that the linguists, the language teachers, and the literati need each other and on the admirable principle that specialists investigating different aspects of a single complex (language, linguistics, culture) might profit by a reduction of barriers and the discovery, after some comparing of notes, that they are all interested in the same problem.

The fourth round-table conference attempts to explore, via panel sessions, three major points: problems of testing in language instruction, technical aids in language teaching and research, and linguistics and the humanities. The problems of testing in language instruction are discussed by John V. Carroll, Harvard, Some Principles of Language Testing; James R. Frith, U. S. Air Force Institute of Technology, Selection for Language Training by a Trial Course; William J. Morgan, Aptitude Associates, A Clinical Approach to Foreign Language Achievement; Paul Thibault, Princeton, Implications of Experience with College Board Language Tests, and Robert Lado, Michigan, Test the Language. The subject of technical aids in language teaching and research is treated by Franklin S. Cooper, Haskins Laboratories, Some Instrumental Aids to Research on Speech; Kenneth N. Stevens, MIT, Synthesis of Speech Sounds by an Electromechanical Vocal Tract, and Ruth Hirsch, Georgetown, Mechanical Aids in Language Teaching. The relationship between linguistics and the humanities is considered by A. H. Marckwardt, Michigan, Introductory Remarks; W. Freeman Twaddell, Brown, Grammar as a Humanity; V. B. Spratlin, Howard, The Audio-Oral Factor in Literary Appreciation; A. T. MacAllister, Princeton, Balanced Emphasis on Language and Culture, and Archibald A. Hill, Georgetown, A Sample Literary Analysis. In addition to the panel papers and a summary of the subsequent discussions the volume contains two speeches given at special luncheon meetings: Structure and Linguistic Analysis by Bernard Bloch (Yale) and Language and Literature in College and University by Norman L. Torrey (Columbia). There are appendices giving the program details, the names of people attending the meetings, and an index of speakers participating in the discussions.

The most significant facts, in the opinion of this reviewer, revealed by the fourth

round-table meeting on linguistics and language teaching are, in terms of the spirit of the meeting, somewhat negative and discouraging. It is apparent from a study of the papers and the post-panel discussions that the process of fragmentation which threatens the humanities has already brought about a situation in which many specialists have come to think of their specialties as virtually closed systems. That whole, which the humanities is supposed to represent, is no longer the sum of its integrated parts but a purely accidental and arithmetical total composed of discrete elements which can be lumped together only by arbitrary definition and not because the parts are understandable in terms of a master pattern. As a result of this psychological set or gestalt, and in spite of a great good will in their attempt to break down barriers, a discouragingly large number of the panelists formulated their problem and defined their terms not with their eye upon a complete system of knowledge but with their attention so narrowly focused that other specialists, as revealed by the discussions, were hardly aware that they were all interested in nothing more different than variant formulations of the same problem. Thus, for example, Professor Spratlin, in his discussion of the audio-oral factor in literary appreciation, is concerned (though he does not announce this to the structural linguists, the phoneticians, the language teachers, nor the psychometricians) almost exclusively with the supra-segmental components of speech which, when added to the visual stimuli of the printed page, quite obviously provide more information and more stimulation for the student. His point, then, is that an increase in stimulation and information (the audio aspect) plus the acquisition of the skill necessary to provide this stimulation and information (the oral aspect) produces an increase in action and reaction (the psychological aspect) which can be interpreted as an increase in literary appreciation.

Formulated in this fashion the matter is hardly a subject either for debate or further investigation. The conclusion breeds neither new research nor additional knowledge. There is, nevertheless, much to be done and much which requires the coordinated teamwork of everyone interested in language. The problem, for the linguist, is to describe these supra-segmental components in terms understandable to everyone; for the semantician, to discover which factors carry information and which operate primarily as stimulants; for the psychologist, to study the difference between conditioned response and discriminatory reaction; for the language teacher and engineer, to develop teaching techniques and mechanical devices which will train the foreign language student to handle these components like a native, and, for the professor of literature, to teach the student to recreate from the printed page the total effect of the living situation. In this hierarchy of specialists the professor of literature stands at the top and occupies, by virtue of his rank, the most vulnerable position. Unless he can formulate his problem in terms understandable to the specialists who are trained to discover new information and new techniques, unless he can define his lack of knowledge in terms which will generate new investigations, he is forever condemned to the frustrating role of the leader who cannot tell his cohorts where to go.

Professor Spratlin's formulation of the problem of literary appreciation in terms of too narrow a focus should be contrasted with Professor Hill's attempt to analyze poetry by means of the mechanics of Aristotelian logic and from an all-embracing, 'humanistic' frame of reference, namely, human activity. His primary assumption that 'poetry, being a human activity, must have pattern and design' is a condensed syllogism: human activity has pattern and design, poetry is a human activity; therefore poetry has pattern and design. The critic's job, then, is to find this pattern and design and to describe it.

It should be noted, at once, that selecting

a too comprehensive level of abstraction as a point of departure produces two major fallacies in the syllogism. First, the major premise, as any psychiatrist or statistician will attest, is not valid. Some human activities are random. Second, the minor premise does not provide a correct equation. Neither poetry nor a poem is an activity. Writing poetry is, but the author is not attempting to describe the patterns of the creative process; he is concerned with the analysis of the static end-product of this process, the poem. This entity exhibits three major types of pattern: first, what may be called linguistic pattern—the structure imposed by the language, the organization selected by the poet, etc., second, literary or genre pattern—the rhyme scheme, stanza length, etc., and, third, the content pattern. Professor Hill is focused, in his model analysis, on the content pattern and he professes that he neither knows nor cares whether his method of revealing the pattern is linguistic or literary. This is a dangerous and unfortunate philosophy. It is dangerous because it is predicated on the assumption that criteria can be validated without categorization, without inquiring what they are criteria for. It is unfortunate in that it closes the door on all specialists who might, if they understood the problem and the processes of its resolution, be able to participate in the investigation, to offer advice about methods, and, likely enough, to provide tested techniques for the validation of the criteria used. Lastly, there exists no persuasive evidence to confirm the assumption that there is a method of investigation which can legitimately be labeled literary. What goes under this rubric is a combination of techniques and criteria which, when they are isolated and defined, characterize a large number of modern disciplines and specialties. This may be demonstrated by re-tracing the steps followed by Professor Hill in his analysis of Tennyson's love lyric, a section of the poetic drama, The Princess.

The first criteria are linguistic and are used to segment the poem into manageable units. Specifically, the critera are derived from a simplified version of the parts of speech, namely, entity, action, modification, and linkage. The method is that of the taxonomist and the referents are categorized on a low level of abstraction. A criterion of relevance is now applied with the result that modification, since it is concerned only with the attributes of either entity or action, is considered peripheral, that is, these elements do not affect the basic structural relationships between entity and action and may, therefore, be disregarded. Thus the entity 'font' in the adverbial phrase 'in the porphyry font' is not counted by Professor Hill in his list of 12 basic entities.

The elimination of the modifying elements focuses attention exclusively upon entities and actions which are directly linked to each other: subject-verb, verb-object, verb-goal (the entity linked to a verb by a preposition, etc.). The pattern which these factors exhibit will, however, not be clear or readily statable unless the various factors can be categorized in terms of some common denominator. The achievement of this categorization requires a reformulation of the problem in terms of non-grammatical criteria. The method, however, remains taxonomic. The problem is to discover those levels of abstraction which embrace the maximum number of factors. Since the poem is a love lyric and the only two characters are 'you' and 'I', the first rung of the ladder of abstraction is based on biological criteria: 'you' and 'I' are classified as male and female. 'I' has a bosom! This establishes a fundamental dichotomy. which, by the process of converting entity into attribute, is now up-graded to a high level abstraction labeled maleness and femaleness. This conversion has two convenient advantages: it retains the basic characteristics of the biological dichotomy and, what is more important, permits the use of non-biological criteria in the process of further classification. As a consequence, the lily of the poem, which is biologically hermaphroditic, can now be classified, according to criteria

derived from literary folklore, as female. Similarly, by using criteria appropriated from Freudian psychology all activity may be categorized as symbolic of (having the attributes of) maleness and femaleness. As a result, rest and inactivity = femaleness; activity and motion = maleness. With all pertinent entities and actions now classified under just two rubrics there remain only two things to be settled: the serial order of events and their denouement which, in this case, may be classified biologically as hoped-for consummation or, by up-grading to metaphysics, as the destruction, by fusion, of the basic dichotomy.

A number of points significant in terms of the purpose of the round table meeting should now be apparent. First, Professor Hill is primarily a taxonomist and as such he shares his method with every major discipline. What distinguishes his procedure and marks him as a humanist is not the method but his willingness to classify in terms of the most diverse criteria—grammar, biology, folklore, psychiatry, etc.—and in terms of extremely high levels of abstraction. In a sense, then, the difference between Professor Hill, as a humanist, and Professor Bloch, as a structural linguist, may be expressed in terms of the number and variety of criteria used in the process of classification. The structural linguist attempts to define and to close the system with which he is working by reducing both the number and the kind of criteria to a minimum; the humanist, in contrast, seeks to extend a selected system by searching for criteria which will bring more and more items into it. Both procedures are theoretically defensible and, as a result, the critical factor is the validity of the criteria used.

This point seems to have received almost no overt attention from the panelists and they failed, as a consequence, to observe that a major difference in criteria divided them into two neatly opposed groups. For example, Professor Carroll, as an educator and psychometrician, approaches the problem of language teaching in terms of tests and measurements. His is concerned with the measurement of language aptitude and the testing of achievement. W. J. Morgan, a clinical psychologist, in contrast, asserts categorically that 'there is no such thing as language aptitude.' These two points of view are certainly incompatible and this incompatibility can be traced directly to the criteria used in defining language. Professor Carroll, without stating his criteria, assumes that language is a 'linguistic code' and that in learning a foreign language 'the individual is learning a new linguistic code.' This suggests that the primary process of learning is memorization and that the critical factor is intelligence. Professor Bloch, in sharp contrast, and also without a description of his criteria, asserts that a set of habits, 'their interrelations and mutual dependencies, constitute the language' and he believes that when a child learns a language he forms a set of habits. This formulation suggests that the primary process in learning a new language consists of the acquisition of a new set of habits and this, in turn, involves, as Morgan points out, the capacity of the individual to make new adjustments, his rigidity of thinking, and a great variety of psycho-biological factors which are quite independent of intelligence.

It should now be apparent that one of the fundamental conflicts brought out by the round-table meeting stems from the consideration of speech, on the one hand, as a human activity and of language, on the other, as a formal system of signs, symbols, etc. The educator (J. B. Carroll, Paula Thibault) and the teacher (Ruth Hirsch) are committed to the assumption that language is a code or a system of symbols which can be taught and memorized purely as an intellectual exercise. This leads the educator to think in terms of a special aptitude, the test maker to conclude, unwillingly, 'that people who can do one kind of linguistic task can do other kinds as well,' and the teacher to concentrate on mechanical aids and drill techniques aimed primarily at the memorization of symbols and fixed patterns.

At the opposite end of this range stands the linguist (Bernard Bloch, Robert Lado, Freeman Twadell) and the clinical psychologist (William J. Morgan). Their point of departure is not language, the abstract system, but speech, the human activity, man talking, man meeting changing situations. Somewhere in the gray middle reaches of this range the dedicated humanist, primarily the professor of literature, wanders about in a kind of limbo, fluctuating between rejecting language (the mechanical system) and embracing speech as an integral part of human activity. As a consequence, Professor Hill can be indifferent to the question of whether his method is linguistic or literary and Professor Spratlin can be unaware of the fact that he is not teaching literature but a very special set of human habits which are characteristic of the process of communicating information and emotional attitudes.

It would seem, as a result, that there is great need for a repeat performance of the round-table conference in which attention will be focused, first, on assumptions, second, on the criteria which are supposed to validate these assumptions, third, on the validity of the systematic formulations derived from these criteria and assumptions, and, finally, on whether it is possible to transform one formulation into the other. Unless such a transformation is possible, it is obvious that the dichotomy which was unconsciously established by the present conference has no counterpart in reality and that a large segment of current theory and research is predicated on assumptions which cannot be validated. If this is so—and the present conference seems to suggest that it is—then the sooner this fact is recognized the better.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

ON THE ETYMOLOGY OF SOME ARAWAKAN WORDS FOR THREE

DOUGLAS TAYLOR

In the chapter of his Primitive Culture (3rd. ed., London 1891) on The Art of Counting, Edward B. Tylor remarked that 'when a word has once been taken up to serve as a numeral, and is thenceforth wanted as a mere symbol, it becomes the interest of the language to allow it to break down into an apparent nonsense-word, from which all traces of original etymology have disappeared'; and he therefore considered it unprofitable to seek 'the origin of numerals not named with reference to hands and feet, and especially of the numerals below five, to which such a derivation is hardly appropriate'. This seems to be borne out by the numeral system of modern Lokono, in which aba *one*, bian ~ biani-*two*, kabun ~ kabuni-*three*, and bibiti ~ biti ~ bis-*four* most probably have no other meanings for native speakers of today; although aba dakabo *one my-hand* or *five*, bian-kotibena *twelve* (cf. -koti *foot*), and loko *person* or *twenty* are still transparent.

Moreover, this 'breaking-down' process may be observed by comparing different stages of the same language or dialect, and different languages of the same family. So, for Lokono, de Goeje (1928) recorded biama *two* and kabuin *three*, while Daniel Brinton (1871) gave kabuhin *three*, bibuti (as well as bibiti) *four*, and aba kutihibena *eleven*; with which compare the forms cited in the preceding paragraph from Hickerson (1953).[1] The only surviving dialect of related, so-called Island Carib[2] has borrowed terms from French for all numerals above three; but the recently extinct (circa 1920) dialect of Dominica had ábą ~ ábana *one*, bíama *two*, írua *three*, and biąburi *four*; this last being transparently derived from the word for *two*, and containing a productive suffix

[1] References are to: Daniel Brinton, The Arawack Language of Guiana, APS-T, 14.427–444 (1871); C. H. de Goeje, The Arawak Language of Guiana (Amsterdam, 1928); Nancy P. Hickerson, Ethnolinguistic Notes from Lokono Lexicons, IJAL 19.181–190 (1953).

[2] 'It is better for science to accept a faulty name which has the merit of existence, than to burthen it with a faultless newly invented one' T. H. Huxley, in Methods and Results of Ethnology (1865).

-buri *multiple*. If further proof of this derivation were needed, it would be found in the recorded numerals of related though long extinct Taino: hequeti *one*, yamoca *two*, canocum *three*, and yamoncobre *four*, (so spelt in the sources).

It will be seen that the Taino word for *one* and the Island Carib word for *three* are unrelated to their Lokono equivalents; the former having obvious Lokono and Island Carib cognates meaning *alone or only*, while the latter (early Dominican élua, recent Dominican írua, Central American órua) was undoubtedly borrowed from Cariban Kalina (oroa, awrua, etc. in modern Guianese dialects). Such ambiguity as may result from the use of a single term meaning *one* or *alone* has often led to supplementation and semantic narrowing; but by the middle of the seventeenth century, Island Carib apparently lacked any obvious cognate of the Lokono word for *three*.

Taken by itself, the resemblance between Lokono -kabo (Brinton has akkabu) *hand* and kabuhin > kabuin > kabun *three* might well be passed over as a coincidence; but it becomes significant on comparison with equivalent forms of related languages, such as Goajiro ahap·u *hand* and apuni *three*, Arauá -safá *hand* and arisafa-há *three*, Paumari -θee-i *hand* and bakuθe-ki *three*; the more so since Kulina -baku *heart, chest, palm (of the hand), sole (of the foot)* suggests that the first element in bakuθe-ki *three* of closely related Paumari may well be a cognate of like meaning.

Now according to Raymond Breton, the word 'eúkê *noyau, & tout autre chose qui a une enveloppe*' was employed by the Indians of 17th-century Dominica to designate, according to qualification or other context, such various things as: *semen, white (of egg), pus (from a boil), starch (of manioc), (iron) rust, 'gold, silver, talcum & other precious things that the earth produces and contains in its bowels', lining (of a basket), centre (of a target), glans (of the penis), vulva*.[3] These last two meanings of (phonemic) óko were also recorded by me from living, Creole-speaking Dominican Caribs; while in the modern dialect of Central America, this word has split into ógo *vulva* and óho *pus, starch (of manioc), extract or essence (of anything)*. Clearly cognate with these forms is (Hickerson's) Lokono -oko *pus, sap*, which is, in all probability, the same morpheme as (Brinton's) Lokono ükkü *heart*, whence his ükkürahü *pus, those of one blood—family, tribe*, and üküahü *person*. It is this word, oko or uku or its parent-form, that I propose as the second member of a one-time Arawak compound meaning *hand-nucleus*, or *three*.

Semantically, the plausibility of this interpretation is enhanced by a piece of information recorded by Gumilla and cited by Tylor (p. 245); 'the mode of showing the numbers with the fingers differs in each nation', he says. 'The Otomacs to say "three" unite the thumb, forefinger, and middle finger, . . . The Tamanacs show the little finger, the ring finger, and the middle finger, . . . The Maipures, lastly, raise the fore, middle, and ring fingers, . . .' Of these, only the last-named speak, or spoke, an Arawakan language.

Phonetically, it is at least not invalidated by insuperable difficulties. If Lokono kabuhin and Taino canocum (both meaning *three*) are cognates, as seems most probable, the latter form must result from 'distance-metathesis' in an older *kamokun;—unless, indeed, it was this hypothetical word whose nasal consonants became transposed through faulty hearing, transcription, copying, or printing. In any case, the final vowel of kabo- *hand*, or the first vowel of -oko (or -uku) *nucleus, etc.*, would regularly be lost in compounding; while according to both Brinton and Hickerson, -n is a subordinating nominal suffix in Lokono. The passage from k to h is common in Arawakan

[3] See Raymond Breton, Dictionaire Caraibe-François . . . , (Auxerre, 1665); édition fac-simile Jules Platzmann (Leipzig, 1892), p. 224 et seq.

languages; so, early Dominican iáka *here* corresponds to iáha *here* of both recent Dominican Island Carib and of Lokono; while Brinton wrote kassaku and kassahu, both meaning *firmament*, on the same line. Nor is the passage from o (or u) to i—or vice versa—unusual when, as is the case in Island Carib and—to judge by both Brinton's and de Goeje's employment of the umlaut—also in Lokono, the former phoneme represents a back unrounded (or it may be, in Lokono, a front rounded) vowel; so, Lokono bibuti > bibiti *four*, and Lokono: Island Carib isi : icígo ~ icógo *head*, isiroko : ógorogo (Dominican ékriko) *flesh*, aduka ~ adoka ~ adika : aríha (early Dom. aríka, recent Dom. aríha) *see*.

Dominica, British West Indies

On Language and Culture

John Lotz

'Ideal targets' for records in sports provide a case-study in the currently fashionable discussion on the interrelation between language and culture. The difference between Anglo-Saxon countries and countries using the metric system may further elucidate the role of the linguistic factor.

In May 1954, when Bannister ran the 'dream mile' (one mile in less than 4 minutes), the newspapers listed a number of other ideal objectives in sports: 9 seconds for the hundred-yard dash, 7 feet for the high-jump, 16 feet for the pole-vault, 60 feet for the shot-put (this one was achieved a few days later). These objectives were for male athletes; one can also add the 'dream mile' for women: one mile in less than 5 minutes (also reached in the same month).

All such 'ideals' aim at lowering the record time for running certain standard distances to a fixed number of seconds or minutes, or, in field events, at raising the length or height of the existing record to a fixed number of length units. The numbers involved are simple integers when the target number is low, and 'round figures' when the number is higher (4 minutes, 12 feet, 60 feet, but not 3¾ minutes for the mile or 61 feet for the shot-put). Such targets of course exist in other sports as well (e.g. in weight-lifting, or the 0.300 batting average in baseball).

That these targets depend on the use of language and not on other factors is demonstrated by the difference between the formulations of such targets in Anglo-Saxon countries and in countries using the metric system. The latter list as ideal targets: 10 seconds for 100 meters, 20 kilometers for the one-hour race, 80 meters for javelin-throwing, etc. Because of this dependence on language, such targets cannot be 'adequately' translated: to try to run 1603 meters in 4 minutes is not the same thing as shooting for the 4-minute mile.

These targets must be within reasonable reach, just above the present records, and therefore an ideal target can normally be formulated in only one measuring system; for instance in the high jump the target in Anglo-Saxon countries is 7 feet, half an inch above the present world record of 212 centimeters: 220, or even 225, is neither 'round' enough, nor realistic enough.

When such an ideal target is fashionable in the world of sports, as e.g., the 4-minute mile, the number of occasions on which that event is put on the program also increases (e.g., the one-mile race was often run in the 'metric' countries of Europe). When the ideal is achieved, that target gradually becomes obsolete; this has happened to the 30-minute 5 kilometers, and it will happen to the dream mile.

It seems clear that these ideal targets for sports records presuppose a general cultural setting of a non-verbal sort: appreciation of number and quantity, high valuation and meticulous recording of top physical performance, a realistic appraisal of man's physical abilities using the present top achievement as a reference point and extrapolating from there. But these targets must also be formulated verbally. The verbal expression is an attributive phrase consisting

of a number followed by measures of time, length, and weight. These expressions of the ideal targets readily lend themselves to analysis and interpretation by reason of their morphemic simplicity and semantic-conceptual transparency.

The second part of the verbal expression, the measure of time and length, is determined by tradition and convention. The foot-yard system, because of the larger number of basic units, allows a much greater variety of ideal targets than the metric system. The expressions for these measuring units are either single morphemes (second, minute, hour, inch, foot, year, mile, meter) or, in the metric system, consistently formed complex morphemes with a unifying stress pattern (centimeter, kilometer, etc.).

The number part of the phrase poses two linguistic problems with psychological and cultural implications: (a) the low, fundamental numbers, and (b) the round numbers.

The beginning of the integer series is said to be the cultural possession of all mankind, although in some cultures only the very first integers occur. They are single morphemes in English and in many languages. The round numbers, on the other hand, imply the selection of certain numbers, the base numbers, from the sequence of the integers (e.g., ten, a thousand, the '-illion' numbers, etc.), which make the manipulation of large numbers possible. The base numbers vary in different cultures (or they may be missing altogether); they may also vary within the same culture for different objects (e.g. in the Western culture twelve is used for eggs and paper; tjog (*twenty*) is used in Sweden for crayfish; the Akkadian 12 and 60 survive in time measures; in general use, the decimal system prevails in pure numbers, but not in measurements in Anglo-Saxon countries). Round numbers are either base numbers alone, or in simple combinations with the fundamental numbers. There are various degrees of 'roundness': any fundamental number in combination with any base number yields a round number (60, 200,000, etc.); certain repetitive types, such as 110, 880, or the 666 of the Apocalyptic Beast, are also regarded as round, as are 5, 50, etc., and 25, 250, etc., in the decimal system. The feeling of what is round may vary in different countries, e.g. 16 is felt to be more round in Anglo-Saxon countries than in Continental Europe because it is a base in measurement—16 ounces to the pound. In all these numerical expressions the purely linguistic part is morphemically simple and the expression consists of a few differing morphemes: this is psychologically axiomatic; the reference, on the other hand, is determined by historical-cultural accident. The status of expressions like quatre-vingt-dix *four score and ten*, and problems such as that of notation and its influence on the verbal expression are beyond the scope of these notes. (The counterpart of the round number is the simple fraction, especially binary divisions like half, quarter, etc., and in the decimal system the fractions ending in .5 or .25 etc., the decimal equivalents of successive halvings, which accounts for the feeling of roundness associated with 5, 25, etc. mentioned above).

The fundamental and the round numbers play a prominent part on the 'desiderative-imperative' aspects of our culture: in law, age limits, length of fish permitted to be caught, speed limits (with their implications for the problems of transportation); in sports: distances to be run specified in round numbers (the marathon distance, 26 miles, 385 yards, is a quasi-traditional distance introduced at the first Olympic Games); in social events: wedding anniversaries, college class reunions, bicentennials; in prices and salaries (the American $9.95 price tag is a deliberate avoidance of such numbers, aimed of course at having the customer psychologically class the article as within the range of the next lower 'round' price). In actual measurements, any degree of accuracy may be used that seems reasonable

(running times are generally given in tenths of a second, distances in yards or fractions of miles; in the metric system, jumps and throws are measured only in whole centimeters), but even in actual reporting of numbers, rounding is common: on old tombstones ages are generally given in round tens, with fives next in frequency. Thus, it is the morphemic-conceptual simplicity and not the objective reference that is decisive.

The use of numbers in such targets determines cultural aims and behavior dependent on these. The simplicity of the linguistic expression and the correlated simple conceptual organization seem to account for the setting up of ideal aims and these cannot be explained by non-linguistic references, biological properties, or the like. The simplicity of the linguistic expression seems to be the organizing principle in other semantic fields also (such as colors, kinship, spatial organization, movements, etc.) in which simple morphemes or morpheme-combinations are set against expressions which might yield a more exact description and identification, but are linguistically complicated and cumbersome.

COLUMBIA UNIVERSITY

REJOINDER TO HAMP'S 'COMPONENTIAL RESTATEMENT OF SYLLABLE STRUCTURE IN TRIQUE'

R. A. LONGACRE

Eric Hamp, in a recent article[1] has clarified some confusion in my statements[2] regarding V?V and VhV sequences in Trique and has made a number of stimulating suggestions relative to the phonemic analysis of that language. The fact that I cannot accept the componential analysis that he suggests in no way detracts from my appreciation of his discussion of my materials. In this short note I wish both to acknowledge Hamp's contribution and to express some objections to his suggested restatement.

There is a point of genuine confusion in my published article. I state that certain V?V and VhV sequences with phonetically short vocoids are to be interpreted as unit phonemes in spite of the intrusive consonantal element, but I nowhere in the article clearly state that ? and h are anything but phonemes in their own right. On the contrary such a statement as the following would possibly imply that ? and h have the same status as full phonemes everywhere: 'Checked vowels are either terminally checked (in respect to the vowel itself) or medially checked by ? or h . . .' I did not list V?V and VhV in my phoneme inventory but only the consonants ? and h along with thirteen vowel phonemes. Both oral and nasal vowels were included in the phoneme inventory, since I did not analyze nasalization as a separate phoneme. Nevertheless I do say that V?V and VhV units are to be regarded as unit phonemes and it is this statement that Hamp has taken at its face value: 'The roster of vowel phonemes implied by Longacre's analysis is a large one: six parallel subsets (V, $\tilde{V}$, V?V, VhV, $\tilde{V}$?$\tilde{V}$, $\tilde{V}$h$\tilde{V}$) each comprising up to six phonemes plus ə.' (E. H., 207). It was with some chagrin that I read Hamp's remorseless tracing out of the logical implication of my statement. Its intent was simply to emphasize the unity of the vowel phoneme in V?V and VhV sequences, in spite of the fact that it is actualized phonetically as two vocoids which I transcribe with two vowel symbols. It was not my intent to include ? and h in any sort of complex vowel phonemes.

As a componential analysis, Hamp's restatement of my materials is weakened by his preoccupation with linear sequence. I have assumed that one of the advantages of componential analysis is that it enables the analyst to perform a phonemic analysis which recognizes items in simultaneous or

[1] IJAL 20.206–209. I refer above to this article as E.H.

[2] Acta Linguistica 7.62–78, especially p. 75–76, fn. 2. I refer above to this article as R.L.

vertical occurrence. This is one of the main points made by Hockett in his Componential Analysis of Sierra Popoluca:[3] 'The point of view here assumed is essentially simply that of removing the linearity assumption from among our working principles' (C.H., 258). It was, therefore, with some surprise that I read Hamp's statement that 'though a component may have a broad phonetic domain with respect to the allophones of neighboring phonemes, it must have a definite statable, and hence not arbitrary, sequential position with respect to the neighboring phonemes' (C.H., 207). Such statements recur throughout the article, so that a discourse analysis of the article as a text would probably reveal the pivotal character of such phrases as the following: 'sequential properties of phonemes' (C.H., 207), 'sequential position with respect to the neighboring phonemes' (C.H., 207), 'clear-cut sequential order in their elements' (C.H., 208), 'sequential order' (C.H., 208), 'maximum sequential form of the syllable' (C.H., 209), 'largest occurring sequences' (C.H., 209).

Having thus committed himself to the assumption of linearity, Hamp proceeds to relegate the components of tone and nasality to a 'component-slot' immediately preceding the vowel of the syllable. What I transcribed as tũ³ *blood* he then transcribed as t³~u which he formulaically summarizes as CTNV (in which C = consonant, T = tone, N = nasality and V = vowel). This causes Hamp some difficulty in that he has been unable—on the basis of the data available to him from my article—to specify with assurance whether the order is TN or NT in the 'component slot'. Hamp frankly admits this difficulty (E.H., 208, fn. 4), but proceeds to state that 'perhaps within the component-slot the linear order is vertical, rather than horizontal'. Here, I am convinced, Hamp has encountered a residue where the assumption of linear sequence is not valid. The linearity assumption cannot be salvaged by saying that 'the linear order is vertical' if there is no longer any 'definite statable, and hence not arbitrary, sequential position' of N over T or T over N. Further Trique data not available to Hamp does not enable us to establish a non-arbitrary horizontal sequence of T and N in terms of Hamp's criterion of phonetic domain, for while it is true that 'T has a marked allophonic effect on the preceding C' (E.H., 208, fn. 4), it is also true that N has a similar allophonic effect on C in that the consonants y and w are phonetically nasalized preceding a nasalized vowel.

The failure to establish a non-arbitrary linear ordering of T and N stems from the simple fact that they are simultaneously occurring features—although influencing in various degrees and manners the allophones of surrounding phonemes. With these features must now be grouped the vowel quality itself to comprise a simultaneous bundle of three features, vowel quality, tone, and nasality, where the first two are obligatory and the latter optional. Just as tone and nasality condition allophones of preceding consonants so the vowel quality itself has a similar influence; the phoneme č is slightly more retroflexed before /a/ than before other vowels; /l/ is more fronted and flapped before high vowels.

The question must of course be raised as to the structural status of the three 'features' of our simultaneous bundle. The tones are phonemes, not simply 'components of phonemes'. In my published analysis I do not abstract a phoneme of nasality along with phonemes of vowel quality, but simply treat all vowels, oral and nasalized, as unit phonemes. If this analysis is retained, then nasality is a 'component' in the usual sense of the term. On the other hand, it is entirely feasible to set up a phoneme of nasality and phonemes of vowel quality. This reduces my phoneme inventory by five phonemes. In this case our simultaneous bundle would be a simultaneous occurrence of three phonemes

[3] IJAL 13.258–267 I refer above to this article as C.H.

and the 'componential analysis' would not be a componential analysis at all in the usual sense of the term.

The above considerations bring to the fore a curious fact regarding Hamp's 'componential analysis'. Componential phonemic analysis—at least as the term is used by Hockett and Harris—refers to components of phonemes. Hamp's components are in every case full phonemes—although one has to read closely to realize this—and his componential analysis is an analysis of 'components of the syllable' (E.H., 207). This is admittedly a new use of the term componential analysis and one that is possibly confusing to the reader. What Hamp actually has done is to analyze nasalization as a separate phoneme, abstract phonemes ʔ and h from the complex VʔV and VhV implied by my statement on p. 76, and then describe the distribution of these phonemes in the syllable. His use of the term 'componential analysis' probably stems from the realization that he is, in fact, handling simultaneously occurring features of the sort more successfully handled by that sort of analysis. But, as we have seen, in the name of this very analysis set up to handle simultaneously occurring features as 'chords' or 'bundles' he rejects simultaneous occurrence in favor of linearity!

In terms of simultaneous features the Trique form tū³ *blood* cited above could be transcribed $t\overset{u}{\overset{\sim}{3}}$ which could be formulaically summarized with Hamp's symbols as $CT.\overset{N}{V}$. (My vertical ordering of the elements of the simultaneous bundle is not to be taken as implying any fixed vertical sequence.) Such a notation is, it seems to me, more useful than one that attempts to order the same elements in horizontal linear sequence. It is, however, too cumbersome to use for regular phonemic transcription which follows the linear convention. Nevertheless, within the limitation imposed by linear convention, there is an approximate representation of the simultaneous occurrence of these features in such a notation as that employed in my article, viz. with vowel symbol on the line, with tilde indicating nasalization over the vowel, and with tone mark as numeral raised exponent-wise following the vowel. I have no quarrel with Hamp's desire to write the tone mark preceding the vowel except that I cannot agree with him that the tone phonemes have the structural characteristic of occurring before the vowel. A tone mark written either before or after a vowel symbol is simply a linear makeshift to represent items in simultaneous occurrence. I myself prefer to write the tone mark after the vowel in that it is less confusing to readers accustomed to seeing tone symbolized in that fashion.

I turn now to Hamp's suggestion that CVʔV and CVhV sequences be transcribed CʔV and ChV. This is a somewhat attractive suggestion in view of the evident clumsiness of my own notation. I nevertheless prefer my clumsy notation to Hamp's neater one in that the former seems to conform more closely to the structural realities of Trique, while the latter possibly springs only from Hamp's desire to order phonemes in simple linear sequence. If, as a matter of fact, some phonemes occur simultaneously with other phonemes, and some phonemes are interposed in the center of other phonemes,[4] is

[4] In view of the fact that my objection to Hamp's suggested revision of the phonemic transcription of VʔV and VhV sequences has led me to assert repeatedly that these sequences are instances of one phoneme interrupting another phoneme, I here marshal briefly my arguments on this point: (a) The VʔV and VhV sequences with phonetically short vocoids are monosyllabic rather than disyllabic sequences (R.L., 76–77, fn. 2). (b) As monosyllabic sequences they could theoretically be considered to consist of either (1) two vowel phonemes separated by ʔ or h; or (2) one vowel phoneme interrupted by ʔ or h. (c) Aside from these particular sequences in question only one vowel occurs to the syllable in Trique. The two phonetically short vocoids in VʔV and VhV sequences are always of the same quality and have the same timing as a word-final uninterrupted or unchecked vowel. There is therefore

there anything to be gained by attempting to obscure these 'messy' facts? The phonetic facts with which our analysis must reckon is that ʔ and h do occur as a glottal constriction and a trough in voicing in the center of the vocoids with which they occur. Naturally, we would assume a phonemic ordering that is different from the phonetic ordering if there were clear structural reasons for doing so; but in the absence of such structural reasons I do not believe that the phonemic ordering should be assumed to be different from the phonetic ordering simply because the analyst has made an a priori assumption that phonemes must occur in straight-forward linear sequence. I do not see that Hamp has made a good case for ordering VʔV and VhV as phonemically ʔV and ʔh sequences. Aside from his argument that this analysis has 'forced on us the recognition of a clear-cut sequential order' (E.H., 208), he cites as structural evidence for the validity of his statement only the following: 'Longacre notes (78–9, fn. 1) that after certain C, when V is i or ʌ, the first V of VhV is replaced in whole or in part by lengthening of the C. We now write the V on the side of the ʔ or h on which its typical behaviour predominates, and attribute the allophonic lengthening to h'. (E.H., 208). It seems sufficient to mention here that the allophonic influence is as marked on the vowel i or ʌ as on the preceding consonant and that the lengthening of the consonant seems to be resultant on the loss of the first vowel mora, so that it might be preferable to leave the h interposed in the center of the vowel with which it occurs and whose first mora it devoices and partially eliminates. We have seen, furthermore, in the analysis of Hamp's T and N relative to his V, the difficulties inherent in an attempt to establish a clear cut sequential order for an element in terms of its influence on allophones of neighboring phonemes. It is doubtful whether such phonetic considerations should outweigh the more obvious phonetic fact of the ordering of the segments themselves.

In further support of my contention that these phonetic sequences with ʔ or h interposed between two short vowels are better transcribed as VʔV and VhV than according to Hamp's suggested revision, I cite certain matters of Trique morphophonemics. Such a form as yā^{3}ʔāhā$^{4\text{-}3}$ *god* occurs in shortened form in the phrase yā^{3}ʔā^{4}h kīhī^{3} *god of the mountain*; the up-glide to tone 3 and the second vowel mora occurring after the h are characteristic only of phrase-final position. Notice that if we adopt the suggested transcription it will be necessary to state that there is a metathesis of ʔ in phrase-final versus non-phrase-final position, since (to use the suggested transcription) y^{3}∼a$^{?43}$∼ha occurs in phrase-final, but y^{3}a∼^{4}ah occurs in phrase-medial. This metathesis would have no phonetic reality. The phonetic reality is rather a dropping of the repeated vocoid in one position relative to the other position, not a shift in the position of h itself. Whatever objections have been raised to grammatical considerations

---

some presumption in favor of interpreting these sequences as unit vowel phonemes plus an intrusive glottal consonant. (d) If, on the other hand, we interpret these sequences as consisting of two phonemic vowels (invariably of the same quality) plus intervening glottal consonant, we posit—on the sole basis of these questionable data—a sort of geminate vowel cluster that has no parallel elsewhere in Trique.

I reject here a couple of alternative analyses; (a) that the ʔ and h in such sequences are prosodic phonemes distinct from the respective consonants; (b) that the VʔV and VhV sequences comprise 'close-knit nuclei' (a sort of super-phoneme), but nevertheless contain two vowel phonemes (Pike). In answer to the first (which I followed in an unpublished paper worked out with Garvin in 1949), I simply argue that the intrusive ʔ and h are phonetically all but identical with the ʔ and h occurring in syllable-initial and syllable-final, and there seems to be no good cause for treating the intrusive glottal elements as distinct from the corresponding consonants. In answer to (b) I argue that, while I have no objection to Pike's 'close-knit nucleus' concept as such, I cannot admit that there are two vowel phonemes in the Trique VʔV and VhV sequences in question.

having weight in one's phonemic analysis, the fact remains that linguists do not go out of their way to choose a phonemicization that complicates their morphophonemic statements when an alternative phonemicization is at hand—especially when the latter conforms more closely to the phonetic realities of the language under investigation.

In closing I wish to consider Hamp's final tabular summary of distributional restrictions in Trique syllable structure (E.H., 209b). He sets up three positions within the syllable: (a) final, wherein occur vowel and ʔ or h; (b) medial, the ('component-slot') wherein occur tones, nasality and ʔ or h; (c) initial, wherein occur ʔ or h and as many as two other consonants. There is, therefore, a maximum sequence of two terms in the final (V ± ʔ or h), of four terms in the medial (T ± T ± N ± ʔ or h) and (T ± T ± T ± N), and of three terms in the initial (± ʔ or h ± C ± C). Therefore, if the tabulation is to be significant in terms of limitation of distribution, it must show less than maximum number in some positions. Hamp's tabulation as it stands meets this test, but further data not in my published article and not available to him along with one datum in the article but overlooked by him modify the tabulation sufficiently to lessen somewhat its significance.

Four statements may be read from Hamp's table: (1) When the final contains two terms, then the medial contains at most two terms. (2) When the medial contains four terms, then the initial contains at most one term. (3) When the medial contains three terms, then the initial contains at most two terms. (4) When the medial contains three terms, then the initial contains at most one term. Statement (1) needs to be revised in the light of one word that occurs in my article (R. L., 76, 1.4) but was not taken into account: dã^{3}rã$^{3-4?}$ *all*. This word, which would be transcribed d^{3}ar^{34}∼a? according to Hamp's suggested transcription, has in its last syllable two terms in the final and three terms in the medial, while Hamp's tabulation allows only for two terms in the medial. Statement (2) needs to be revised on the basis of some additional data not in the published paper, viz. gi$^{3?}$yãhã$^{4-3}$ *festival*, which would be transcribed by Hamp as g^{3}i?y^{43}∼ha. This word has in its last syllable four medial terms and two initial terms, while Hamp's tabulation allows only one initial term here. Statement (4) likewise needs to be revised in the light of further data not in my published paper: gwi^{3-5-4} *people* and ?we^{3-4-3} *market-place*. In these words there are three tones in the medial and two terms in the initial, while Hamp's tabulation allows for only one term in that position. Our revision of statements (2) and (4) telescopes statements (2), (3), and (4) which now may be summarized: When the medial contains as many as three terms, then the initial contains at most two terms. Alongside of this we put our revision of the first statement from Hamp's chart: When the final contains two terms, then the medial contains at most three terms. To summarize the two restrictions together we may say that: When the final contains maximum terms, then the medial contains at most one less than maximum terms; when the medial contains as many as one less than maximum terms, then the initial also contains one less than maximum. Thus as Hamp himself suspected 'it is highly likely that an inspection of the complete corpus would show the above statements to be quantitatively wrong in at least some respect' (209). Nevertheless, Hamp's general statement is validated that 'the limitations on the maximum number of terms which may be present simultaneously in the syllabic element operate in a direction from final to medial to initial' (209).

However, Hamp's tabulation even as revised, leaves me somewhat uneasy for two reasons: (a) The restriction of the number of terms in the initial to one less than maximum, i.e. to two, may not be valid in view of the fact that clusters of three consonants are somewhat rare in Trique and further research may turn up a word which would

show this restriction to be illusory. (b) The restriction of the medial to one less than maximum, i.e. to three terms, might better be stated in terms of the particular phonemes and/or components involved (tone, nasality, ? or h). The obligatory component in Hamp's 'component-slot' is T (tone). Two terms in the component-slot consist of T plus any other component, T, N, ? or h. Three terms in the component slot consist of T N ? or h, or of T T plus any other component (T, N, ?/h). Finally four terms in the component-slot consist of TTN?/h or TTTN. The pertinent restrictions of one vs. two terms (i.e. V plus ?/h) in the final are: (1) no more than two tones occur in medial, when there are two terms in the final, but three tones may occur in the medial when there is but one term in the final. (2) ?/h do not occur in the medial when there are two terms in the final, i.e. ?/h do not occur both in medial and final of the same syllable.

The mutual relations of vowel quality, ? and h, tones and nasality would be summarized as follows in terms of my own analysis: (a) One and only one vowel quality occurs to a syllable. (b) The vowel of a syllable may be open; closed by ? or h; or interrupted by ? or h. The vowel does not occur both closed and interrupted. (c) Only one or two tones occur on interrupted and closed vowels. (c) Three tones occur only on open vowels. (e) Any number of permitted tones may be accompanied by nasal timbre on the vowel.

My objections to Hamp's 'componential restatement' of Trique syllable structure may be summarized as follows: (1) It attempts to force into linear sequence features that occur as simultaneous bundles thus rejecting one of the chief advantages of componential analysis in general. (2) It likewise attempts to force into straightforward linear order sequences in which one phoneme occurs interposed in another phoneme. (3) It is not a componential analysis at all in the usual use of the term in that the 'components' are in each case phonemes, and that the article is essentially a distributional statement of these phonemes in regard to the syllable matrix. (4) As a distributional statement, this sort of 'componential analysis' does not seem to me to be as pertinent as a more traditional type of statement such as that given in the above paragraph.

SUMMER INSTITUTE OF LINGUISTICS

THE WALAM OLUM PROJECT

JOHN WITTHOFT

Concurrent with publication of the recently issued Walam Olum report of the Indiana Historical Society, and as the next step in the Walam Olum research project, C. F. Voegelin and John Witthoft have undertaken a supplemental linguistic study of the Walam Olum text, based on the information in older published and manuscript sources on the Delaware language, and the reinterpretation of the corpus of Delaware data in these sources through modern spoken Delaware. While such a study is felt to be of lesser importance than that through spoken Delaware sources and thus holds a subsidiary position, it is still of great consequence to Walam Olum studies. Employment of this method in complete independence of the method used in the recently published report gives us another check on the meaning and content of the Delaware text, and we feel that the two independent procedures as applied to the same text, with subsequent correlation and reanalysis on still other levels, represent the greatest refinement we can as yet bring to the study and interpretation of this important text. We expect that we may impose on various of our colleagues for suggestions and criticism in the very near future when we have fair copies of our present manuscript prepared, and during the later stages when we attempt to carry interpretation into the field of comparative Algonkian studies.

PENNSYLVANIA HISTORICAL AND
MUSEUM COMMISSION, HARRISBURG, PA.

ON THE STATE AND TASKS OF SOVIET LINGUISTICS. [In Russian.] By V. V. Vinogradov and B. A. Serebrennikov. Izvestija Akademii Nauk USSR, Otdelenie literatury i jazyka 12.317 ff., 1954.

EDWARD STANKIEWICZ

This article, by two leading Soviet linguists, contains a summary of the achievements of Soviet linguistics in the last four years and a program of its future tasks.

It is based on reports delivered earlier by these linguists during the sessions of the Learned Council of the Institute of Linguistics of the Academy of Sciences.[1] The publication of these reports in an extended version and as the joint effort of the two Soviet Academicians coincides with the fourth anniversary of the liquidation of Marr's school and of the new course of Soviet linguistics which followed the personal intervention of Stalin in the 1950 linguistic controversy. As a brief survey of the whole field of Soviet linguistics it is of interest to Western scholars who have watched the development of Soviet linguistic science with justified curiosity.[2]

The period of Soviet linguistics since 1950 is viewed by the authors as 'the revolutionary epoch' of its development. The break with Marrism was of the greatest significance in that it enabled the Soviet linguists to define the subject-matter of their discipline and to separate it from other sciences, such as sociology, ethnology, history, archeology, etc.

The article lists the basic theoretical problems which face Soviet linguistics: the study of grammatical structure and of the basic lexical fund, the question of language as a social phenomenon, the internal laws of language development, problems of lexicology and of historical semantics, the study of dialects and of the formation of national languages and the study of styles and of literary languages.

The order of importance and the treatment given to these problems in Soviet linguistic literature clearly correspond to the significance assigned to them in Stalin's linguistic pronouncements during the 1950 controversy.[3] The articles and discussions of the last four years on linguistic principles and methodology hinge mainly around the formulations given to them by Stalin and do not generally go beyond an exegesis of his statements raised to official doctrine, although the interpretations of the Soviet linguists of particular problems are not always homogeneous.

According to the authors, debates and studies on general theoretical questions were unavoidable in the period immediately after the liquidation of the Marrist School, a period which they now consider past. The results of the discussions and of the broad, theoretical investigations were on the whole unproductive, because they suffered, in their opinion, from an excess of speculative, abstract formulas; an 'overdose' of theory also marks the most recent collective works prepared by the Institute of Linguistics: Problems of Grammatical Structure, and Problems of Lexicology. It is therefore imperative, according to them, to devote more attention to concrete problems, and to the collection of new empirical material. This kind of work and orientation should mark the new or second phase of Soviet linguistics in its post-Marrist development.[4]

[1] Voprosy Jazykoznanija 3.151, 1954.

[2] See The Soviet Linguistic Controversy (Columbia Univ. Press, 1951); H. Rubenstein The Recent Conflict in Soviet Linguistics, Lg 27.281–287, 1951; R. L'Hermitte Les problèmes des lois internes de développement du langage et la linguistique Soviétique, Word 10.189–196, 1954.

[3] J. V. Stalin, Marksizm i voprosy jazykoznanija. Moscow, 1954. It is interesting however that the article under review uses some well known phrases from Stalin's writings without quotes and makes explicit reference to his works only on one occasion.

[4] In the original reports the first period dates from 1950 to 1952, when the current period presumably began. Cf. Voprosy Jazykoznanija 3.151 ff, 1954.

The primary task of Soviet linguistics is to study the history of various languages. The leading role in diachronic linguistics belongs to the field of Indo-European, which provides the methods for other areas of comparative-historical research.

However, according to the authors, Soviet linguistics cannot boast of new, original contributions to the study of IE languages. Soviet scholars adopt uncritically the theories of Western scholars (like Benveniste and Kuryłowicz). The authors do not indicate any new techniques or principles which should guide 'Marxist' comparative-historical research, except for a greater emphasis put on external standards to which the history of IE languages should be referred. A greater stress should be placed on the study of history of the peoples 'who were the bearers of the languages'[5] and more importance should be ascribed to lexical and etymological problems.

The study of the IE languages should have as its chief objective the reconstruction of their older stages and of the Ursprache in its main contours.[6] The authors seem to differ in this respect from other Soviet linguists who consider the comparative method a means to the discovery of 'internal laws of development' (postulated by Stalin as 'the chief task of linguistics'[7]), and the reconstruction of the protolanguage a secondary and abstract problem.[8]

Very slow progress is being made, according to the authors, in applying the comparative method to the study of Finno-Ugric and of Turkic languages. Skepticism as to the validity of its application reigns particularly among Soviet Turkologists, who question the genetic unity of the Turkic languages and the possibility of reconstructing the protolanguage. The authors point to the pioneering works of Radlov, Ašmarin and Bogorodickij, as guides in historical studies of the Turkic languages. The authors make no reference to theories of relationship through geographical proximity, current among scholars of the Turkic languages, their interest being solely in genetic problems of affinity.

Also bad, according to the authors, is the situation of dialect studies in the Soviet Union. Work on dialectological atlases and monographs is not coordinated and general principles of research are lacking. Dialect studies are urgently needed particularly in those areas where Russian has supplanted the local dialects (e.g. Southern Karelian, Vepsa and some Mordvinian dialects).

Greater emphasis should also be put, according to them, on problems of prehistory and ethnology and of their relation to linguistic problems. Residues of 'paleontological' (i.e. Marrist) theories are still lingering, according to them, among Soviet linguists. There is no clarity on the distinction between ethnic and linguistic origins, and such theories as the origin of Azerbaijani from Albanian or Median, or the 'Hittite-Iberian theory' (i.e. the theory of the genetic relationship between Near Eastern, Caucasian and Mediterranean languages[9]) are still advanced by some linguists.

The textbooks on the history of particular languages are, in the opinion of Vinogradov and Serebrennikov, likewise poor in methodology and in presentation. Even the works on comparative Slavic grammar and on the history of Russian are unsuccessful. The Historical Grammar of the Russian

[5] The methodological requirement to study the history of languages in relation to the history of peoples was set forth by Stalin in the pamphlet already cited (fn. 3).

[6] Trubetzkoy's hypothesis of an original plurality of IE dialects (Acta Linguistica I, 1939) is rejected by the majority of Soviet linguists as a concept typical of 'anti-historical structuralism'. Cf. Voprosy Jazykoznanija 3.102, 1954.

[7] Stalin, op. cit. 33–34.

[8] Voprosy grammatičeskogo stroja i slovarnogo sostava jazyka. Trudy Instituta Jazykoznanija 1.21, 1952; Voprosy Jazykoznanija 1.163, 1954.

[9] For a review of the latter theory see Čikobava, Vvedenie v jazykoznanie, v. I, 227, 1952.

Language by Černyx, and the Historical Morphology of the Russian Language by Kuznecov follow the old, i.e. neogrammarian, tradition. The History of the Armenian language by Ačarjan is only a historical study of Armenian loanwords; the Historical Grammar of Erza by Bubrix is a collection of various essays on Erza Mordwin; the History of Scandinavian languages by Steblin-Kamenskij treats the history of various Scandinavian languages without correlation. It is worth mentioning that the last book received favorable reviews[10] by some Soviet linguists for its attempt to combine linguistic and extralinguistic history. The innovation lies, actually, in providing an outline of the history of the Scandinavian peoples in addition to the history of the languages, which is supposed to prove the interrelation between language and society.

Textbooks on the history of various languages of the Soviet Union are in preparation. However, the authors of the article express doubt whether these studies will be successful, as long as there are 'no hard decisions on many basic theoretical problems of development of literary languages.' That the lack of general theoretical principles and techniques worries the Soviet linguists is indicated, for instance, by the fact that some Soviet linguists, (preferring to play it safe in problems of theory) proposed that the editors of Voprosy jazykoznanija provide an outline of general linguistic concepts which could serve as a guiding rod in their empirical research.[11]

The article under review makes no secret of the fact that the linguistic institutes of the Soviet Union are passive in regard to controversial, theoretical problems. Discussions organized with the view of tackling general questions end in failure or in practical, pedagogical debates on teaching and translating from Russian into the national languages (for example, the six month discussions in the Chuvash newspaper Kommunizm jalave on the Chuvash case-system). The authors claim better accomplishments in the field of descriptive grammars. The descriptive grammar of contemporary Standard Russian is considered one of the most successful achievements of Soviet linguistics.[12]

It is appropriate to remark that although very thorough and exhaustive, this grammar by no means meets the standards of a structural linguist. It abounds in many redundant and ill-defined rules and does not always make a distinction between the spoken and written language, let alone a description in terms of relevancy and of patterns.

The compilation of comparative grammars of Russian and other national languages, and of bilingual dictionaries pursues primarily practical purposes and lacks again, according to the authors, unified methods and a unified terminology. In the field of lexicology, the Soviet linguists are faced with the problem of translating the new Soviet terms into the national languages (i.e. the introduction of calques) or with the transfer of the Russian terms into these languages, (i.e. the introduction of loanwords). It is apparent from the reports in Soviet linguistic journals that the latter solution is chosen.[13] Although the Soviet Union is, in the opinion of the authors, the

[10] Voprosy Jazykoznanija 3.126, 1954.

[11] This proposition met with protest on the part of one of the editors; see Voprosy Jazykoznanija 3.151, 1954.

[12] Grammatika Russkogo Jazyka. Ed. by Acad. of Science of USSR, 1952. The second volume, Sintaksis Russkogo Jazyka, is to appear at the end of 1954.

[13] See P. P. Baraškov, Po povodu vyskazyvanij A. E. Mordinova po voprosam jakutskoj orfografii, Izv. Ak. Nauk 13.397–400, 1954. The attempts at Russification are by no means limited to languages without a literary tradition; significant in this respect is the 'cultural value' assigned to Russian elements in the 'enrichment' of such literary languages as Ukrainian. See J. A. Bagmut, Rol' perevodov proizvedenij klassikov marksizma-leninizma v obogaščenii ukrainskogo jazyka. Izv. Ak. Nauk 13.330–338, 1954.

world's leading country in the publication of bilingual dictionaries, it has not produced theoretical studies on lexicography corresponding to the practical achievements in this area.

The general lexicographical works of importance remain the late Ščerba's studies Opyt obščej teorii leksikografii and his introduction to the Russian-French Dictionary. A general Marxist theory of lexicology, lexicography and semantics is still a problem for the future. The recent study by Bulaxovskij on these problems is not mentioned in the article, evidently because it was considered a failure by Soviet linguists.[14]

The article lists finally the problems which were the subject of various discussions in the Institute of Linguistics and in the leading linguistic journals. It considers them on the whole unsuccessful. These discussions centered around the problems of the 'internal laws of linguistic development',[15] around the concept of the phoneme,[16] and around problems of stylistics and of literary languages. The problem (raised in Stalin's writings) of the relation of language and thought has not been tackled either by linguists or by philosophers, for each expects solutions from the other.

The article blames the Soviet linguists for embracing 'bourgeois' theories without an attempt at interpreting them in the light of Marxist philosophy, and for asserting on the other hand their Marxism and originality by the repudiation of all Western linguistic theory as bourgeois and idealistic.[17]

To this reviewer the article is of particular interest because of its moderate tone, its objective evaluation of the achievements, trends and failures of Soviet linguistics and as an indication of the state of a science which is struggling for its development within the limitations of a dogma, after having liberated itself from a dogmatism which cramped its growth for more than three decades.

INDIANA UNIVERSITY

RECENT PUBLICATIONS

General Linguistics

Bar-Hillel, Yehoshua. Logical Syntax and Semantics. Lg 30.230–237, 1954.

de Felice, Emidio. La Terminilogia Linguistica di G. I. Ascola e della sua Scuola. Spectrum (Utrecht, Anvers): Comité International de Linguistes, 1954. 36 pp.

Garvin, Paul. Delimitation of Syntactic Units. Lg 30.345–348, 1954.

Guiraud, Pierre. Bibliographie Critique de la Statistique Linguistique. Spectrum (Utrecht, Anvers): Comité International Permanent de Linguistes, 1954. xix + 123 pp.

Hamp, Eric P. A Note on Structural Complexity. SIL 12.1–7, 1954.

Linguistics Today. Word, vol. 10, nos. 2–3, 1954. Published on the Occasion of the Columbia University Bicentennial. [Articles: Martinet, André. The Unity of Linguistics. Pp. 121–125. Bazell, C. E. The Choice of Criteria in Structural Linguistics. Pp. 126–135. Frei, Henri. Critères de delimitation. Pp.

[14] M. Aleksandrov's review of Bulaxovskij's Vvedenie v jazykoznanije, in Voprosy Jazykoznanija 1.162, 1954.

[15] Cf. R. L'Hermitte, op. cit. (fn. 2).

[16] Cf. Izv. Ak. Nauk 11, nos. 4, 5, 6, 1952; 12, no. 4, 1954.

[17] Particularly violent in tone are the attacks against structuralism, against American descriptive methods, and against ethnolinguistics. See O. S. Axmanova, Glossematika Lui El'msleva kak projavlenie upadka sovremennogo buržuaznogo jazykoznanija. Voprosy Jazykoznanija 3.25–47, 1953; M. M. Guxman, Protiv idealizma i reakcii v sovremennom amerikanskom jazykoznanii. Izv. Ak. Nauk 11.281–294, 1952; M. M. Guxman, Ė. Sapir i ètnografičeskaja lingvistika. Voprosy Jazykoznanija 1.110–127, 1954. The level of these articles, which identify Sapir with racism, Hjelmslev and Bloomfield with reactionary philosophy and imperialism, and which tendentiously ignore the variety of orientations among American linguists, is beneath discussion.

136–145. Harris, Zellig S. Distributional Structure. Pp. 146–162. Hjelmslev, Louis. La stratification du langage. Pp. 163–188. L'Hermitte, R. Les problèmes des lois internes de développement du langage et la linguistique soviétique. Pp. 189–196. Halle, Morris. The Strategy of Phonemics. Pp. 197–209. Hockett, Charles. Two Models of Grammatical Description. Pp. 210–234. Wells, Rulon. Meaning and Use. Pp. 235–250. Benveniste, E. Problèmes sémantiques de la reconstruction. Pp. 251–264. Malkiel, Yakov. Etymology and the Structure of Word Families. Pp. 265–274. Borgström, Carl Hj. Internal Reconstruction of Pre-Indo-European Word-Forms. Pp. 275–287. von Wartburg, Walther. Le *Französisches Etymologisches Wörterbuch*: évolution et problèmes actuels. Pp. 288–305. Swadesh, Morris, Perspectives and Problems of Amerindian Comparative Linguistics. Pp. 306–332. Tovar, Antonio. Linguistics and Prehistory. Pp. 333–350. Haudricourt, André G. Comment reconstruire le chinois archaïque. Pp. 351–364. Vogt, Hans. Contact of Languages. Pp. 365–374. Bottiglioni, Gino. Linguistic Geography: Achievements, Methods, and Orientations. Pp. 375–387. Weinreich, Uriel. Is a Structural Dialectology Possible? Pp. 388–400.]

Malkiel, Yakov. Language History and Historical Linguistics. Rom. Phil. 7.65–76, 1954.

Malkiel, Yakov. The Place of Etymology in Linguistic Research. Bull. of Hispanic Studies 31.78–90, 1954.

Sommerfelt, Alf. The Origin of Language, Theories and Hypotheses. Jour. of World History (Paris) 1.885–902, 1954.

American Languages

Bibliographical Bulletin of American Anthropology. Vols. XV and XVI, 1952–53. Mexico, D. F.: Inst. Panamericano de Geografia e Historia, 1954. Part One, 484 pp.; Part Two, 373 pp.

Caudmont, Jean. Materiales para el estudio lexicográfico de la lengua Inga. Divulgaciones Etnologicas (Colombia) 3.165–186, 1954.

Dumézil, Georges and G. Andres Alencastre. Fêtes et usages des Indiens de Langui (Dept. du Cuzco). JSAP 41.1–118, 1953.

Gerhard, Robert H. General American Pronunciation. A Phonetic Reader for Japanese Students. Tokyo: Sanseido Press, 1954. iv + 226 pp.

Groot, A. Willem de. Scientific Grammar of Present-Day English. Vancouver: Univ. of British Columbia, 1954. Mimeo. vi + 117 pp.

Hirsch, David I. Glottochronology and Eskimo and Eskimo-Aleut Prehistory. AA 56.825–838, 1954.

Huber, Konrad. Contribution à la langue mučik [Peru]. JSAP 42.127–134, 1953.

Partridge, Eric. The Concise Usage and Abusage. A Short Guide to Good English. New York: Phil. Lib., 1954. ix + 219 pp.

Rivet, Paul. La langue mašubi. JSAP 42.119–125, 1953.

Rivet, Paul and Robert de Wavrin. Les Nonuya et les Okáina [Amazon]. JSAP 42.333–390, 1953.

Publications Received

America Indigena, v. 14, no. 4, 1954. Mexico, D. F.

American Speech, v. 29, no. 3, 1954. New York: Columbia University.

Antropologia e Historia de Guatemala, v. 5, no. 1, 1953.

Boletin Indigenista, v. 14, no. 3, 1954. Mexico, D. F.

Divulgaciones Etnologicas, v. 3, no. 5, 1954. Colombia.

Japanese Journal of Ethnology, v. 12, no. 3, 1953.

Journal de la Société des Américanistes, v. 42, 1953. Paris.

Journal de la Société des Océanistes, v. 9, no. 9, 1953. Paris.

Language, Journal of the Linguistic Society of America, v. 30, nos. 2, 3, 1954.

Revista do Museu Paulista, v. 7, 8, 1953–54. São Paulo, Brazil.

Ricerche Linguistiche, v. 3, 1954. Rome.

Slovo, v. 3, 1953. Zagreb.

Tribus, Jahrbuch des Linden-Museums, 1952–53. Stuttgart.

Word, Journal of the Linguistic Circle of New York, v. 10, nos. 2, 3, 1954.